Available online at www.sciencedirect.com
ScienceDirect
Journal of Biomedical Informatics 41 (2008) 293–317
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Comparing and consolidating two heuristic metaschemas

Yan Chen [a,b,*], Yehoshua Perl [a], James Geller [a], George Hripcsak [c], Li Zhang [d]

[a] New Jersey Institute of Technology, Newark, NJ 07102, USA
[b] Computer Information Systems Department, Borough of Manhattan Community College, CUNY, New York, NY 10007, USA
[c] Department of Medical Informatics, Columbia University, New York, NY 10032, USA
[d] Business and Computer Science Department, College of The Sequoias, Visalia, CA 93277, USA

Received 5 May 2007
Available online 17 November 2007

Abstract

A metaschema is an Abstraction Network of the UMLS's Semantic Network (SN) obtained from a connected partition into its collections of semantic types. A metaschema can support comprehension, visualization and navigation of the SN. In this paper, a methodology is presented to evaluate and compare two given metaschemas, based on their structural properties. This methodology was applied to the top-down metaschema and the bottom-up metaschema, derived from two studies involving two groups of UMLS experts. By the Cronbach alpha estimate, the bottom-up metaschema is reliable, while the top-down metaschema is not. An algorithm was designed to yield a consolidated metaschema enjoying the best and avoiding the worst of the two given metaschemas. The consolidation algorithm was applied to the above two metaschemas, The resulting consolidated metaschema has better structural properties than either of the two metaschemas and better supports user orientation visualization, and navigation of the Semantic Network.

Keywords: UMLS; Semantic Network; Abstraction Network; Metaschema; Comparison; Consolidation; Evaluation

1. Introduction

The Metathesaurus (META) [1,2] and the Semantic Network (SN) [3–6] are two knowledge resources of the Unified Medical Language System (UMLS) [7–9]. The META is a valuable concept repository containing more than 1,300,000 (2006 AC) biomedical concepts from various source terminologies. The SN provides a high-level abstraction of the biomedical domain and consists of 135 broad categories called semantic types (STs), by which all concepts of the META are categorized. The SN can help with user orientation and navigation in the large META knowledge base and with interaction with the knowledge embedded in the UMLS.

Although the SN provides a high-level abstraction of the META, it is still difficult for a user to obtain full comprehension of the SN, since there are about 7000 non-IS-A semantic relationships connecting pairs of STs. Previous research has been conducted on reducing the complexity of the SN. McCray et al. [10] developed a methodology for aggregating STs into 15 groups based on 6 general principles: semantic validity, parsimony, completeness, exclusivity, naturalness and utility. Kumar et al. [11] used a derivation formalized in predicate logic to reduce the complexity of the SN. We have developed partitioning techniques to obtain a *metaschema* [12], which serves as a compact abstraction of the SN.

According to the definition in [12], a metaschema is an abstraction based on an underlying partition of the SN into connected groups of STs. Each group is represented by a single metasemantic type (MST). The purpose of a metaschema is to present a compact abstraction-level network of the SN, where each MST represents a subject area of the SN. That is, the STs of each subject area constitute a group in the partition underlying the metaschema. Similar to the SN itself, a metaschema of the SN is formally a

* Corresponding author. Address: New Jersey Institute of Technology, Newark, NJ 07102, USA.
E-mail address: yc7@njit.edu (Y. Chen).

1532-0464/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2007.11.001

directed network which consists of a set of nodes, the MSTs, connected via hierarchical *meta-child-of relationships* and semantic *meta-relationships*. For more details see Section 2.1. In [13], we extended the notion of a metaschema to a Directed Acyclic Graph (DAG) network, rather than the tree structure of the SN. We then obtained two different metaschemas from the Enriched Semantic Network [14], which is an extension of the SN having a DAG structure.

We designed algorithms to generate two different metaschemas, the *cohesive metaschema* [12], and the *lexical metaschema* [15]. Each of these two metaschemas can serve as a higher-level abstract view of the SN to help users' understanding of the complex SN. An important assumption underlying the construction of these two metaschemas is that even though they are generated by algorithmic processes, they effectively yield subject areas meaningful and useful to a human. In order to evaluate the validity of this assumption, we conducted a heuristic top-down study [16]. In this study, a group of experts, who published on UMLS research or related subjects, was recruited, with each expert charged to derive his/her own metaschema. A consensus metaschema of all the experts' metaschemas was then derived.

However, the result of the top-down study was disappointing, since the metaschemas obtained by the experts varied widely while we expected a good degree of agreement among experts. As a consequence, we introduce in this paper an alternative bottom-up approach. We report in this paper on a second study of experts applying this approach.

Naturally, one wants to compare the top-down and the bottom-up consensus metaschemas to find out which is better fitting for evaluating the algorithmatically obtained metaschemas. Since the two consensus metaschemas are results of human considerations, only a comparison and evaluation of their structural properties can be objective. In this paper, we present methods for comparing two metaschemas. We will further introduce several measures to help in assessing the quality of a metaschema for supporting user orientation into the SN. These measures are structural measures, intended to reflect the ease of comprehension and orientation. Those structural measures were used for evaluating the two metaschemas. As will be shown, each of them has pros and cons, in terms of its structural properties. We desire a metaschema that can best facilitate user orientation into the SN, by enjoying the advantages of each of the consensus metaschemas and avoiding their disadvantages. To this end, we have developed an algorithm to obtain a consolidated metaschema of the two given metaschemas. The consolidated metaschema obtained can serve as a yardstick for the measurement of the quality of the metaschemas generated by algorithms [12,15], since it is derived from experts' metaschemas. We include a glossary of the different metaschemas mentioned in this paper.

In [17], Gu et al. use the metaschema paradigm to locate concepts with high likelihood of errors. A metaschema provides various compact partial views of the SN to help users with visualization, orientation and navigation in the SN [12].

The metaschema framework can be extended beyond the UMLS to any dual level terminological system which consists of an upper level terminology of broad categories, in addition to the concept repository, with assignment of categories for every concept. Such a terminological system will have advantages in supporting abstraction, navigation and integration. Metaschemas and their consolidation can further support abstraction of the upper level terminology.

One effort in this direction appears with regard to the IEEE Standard Upper Ontology (SUO) [18,19] as an upper level terminology for the WordNet terminology [20]. Assignment of the SUO categories for the WordNet terminology is described in [21].

Another related dual level terminology is suggested in [22] for the Medical Entity Dictionary (MED) [23]. The upper level terminology there is called schema (following the Object-Oriented database paradigm [24]). A partition for this schema is suggested in [25]. This partition can probably serve as a basis for a metaschema of a different kind but further research is needed.

In addition to the notion of metaschema, other previous work has focused on different methods to facilitate UMLS knowledge comprehension and visualization. Bodenreider and McCray described how to use visualization of semantic relationships as important indicators to explore coherence of semantic groups and help in auditing and validating the SN [26]. In [27] Nelson et al. presented the Hypercard browser MetaCard to enable users to extend the browsing process from META to a variety of different knowledge sources. In [28], knowledge exploration tools using levels of indentation to represent items standing in hierarchical relationships were used for displaying biomedical hierarchies in environments such as Protégé-2000. A review of knowledge visualization and navigation in the medical domain was presented by Tuttle et al. in [29].

2. Background

2.1. A metaschema of the SN

The notion of a metaschema was introduced in [12] as an abstraction of the SN. An ST group is called *connected* if its STs together with their respective IS-A links constitute a connected subgraph of the SN hierarchy with a unique root. A partition is called *connected* if each of its ST groups is connected. A metaschema is based on a connected partition of the SN, where the SN's IS-A hierarchy is partitioned into disjoint ST groups. Fig. 1 shows a partition of the **Event**[1] portion of the UMLS SN hierarchy. Each box represents an ST. Each arrow represents an IS-A link.

[1] Semantic types will be written in bold style and MSTs will be written in "small caps" style.

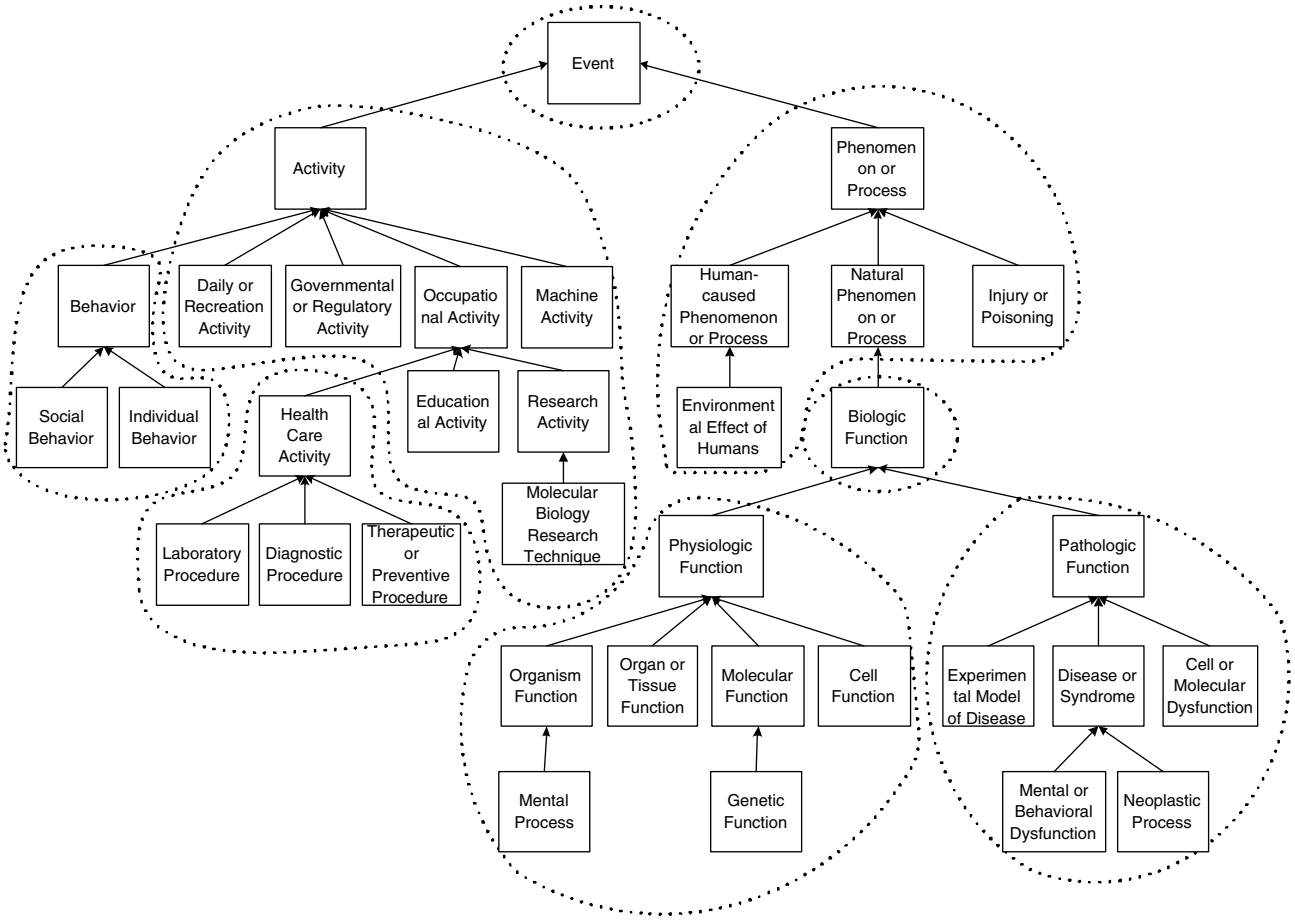


Fig. 1. A connected partition example of the **Event** hierarchy of the SN.

Dotted lines circumscribe groups of STs which are close in meaning to each other. Additionally, while an ST group can be a singleton (i.e., a group of one ST), we require that such an ST cannot be a leaf in the SN's hierarchy. This condition was imposed because the metaschema should manifest some size reduction of the SN, which singletons do not contribute to. However, a singleton containing a non-leaf ST with more than one child is allowed, since it may express an important internal branching point in the metaschema. For example, in Fig. 1, the singleton {**Biologic Function**} serves as a branching point for the groups rooted at **Physiologic Function** and **Pathologic Function**.

In a metaschema, each ST group of the partition is represented by a single node, called a *metasemantic type* (MST) named after the root of the group. Two kinds of relationships, the hierarchical *meta-child-of* relationships and the non-hierarchical *meta-relationships*, connect MSTs. Fig. 2 shows the metaschema hierarchy corresponding to Fig. 1. The number of STs in each MST is listed in parenthesis following its name.

A *meta-child-of* relationship ("*meta-child-of*" for short) is a link between two MSTs representing an IS-A relationship between two STs of the corresponding ST groups. More specifically, let $\mathbf{A}_i$ and $\mathbf{B}_r$ be STs in the ST groups of MSTs A and B, respectively (see Fig. 3). Furthermore,

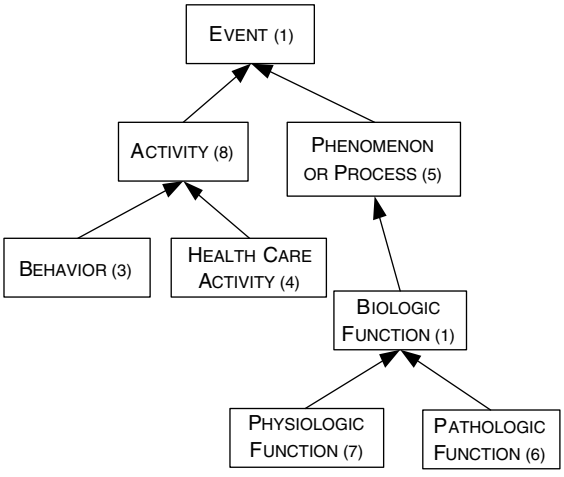


Fig. 2. Metaschema hierarchy corresponding to the partition of the **Event** hierarchy of Fig. 1.

let $\mathbf{B}_r$ be the root of B and let $\mathbf{B}_r$ IS-A $\mathbf{A}_i$. Then in the metaschema, we define a *meta-child-of* directed from B to A. Note that the ST $\mathbf{A}_i$ does not need to be the root of its MST. Only the source $\mathbf{B}_r$ has to be a root in order for a new *meta-child-of* to be induced in the metaschema. A *meta-relationship* is a link between two MSTs representing

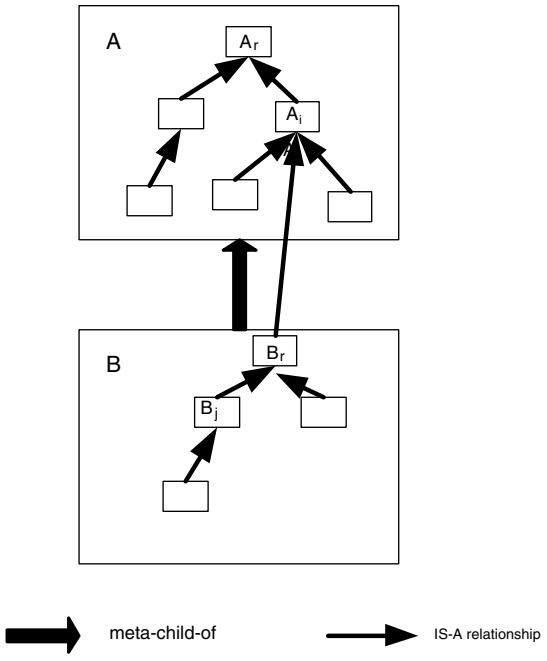


Fig. 3. Interpretation of the definition of *meta-child-of*.

a specific semantic relationship (non-IS-A relationship) between the two corresponding ST groups (for details see [12]). The derivation of the *meta-child-ofs* and the *meta-relationships* is motivated in detail in [12].

For example, the hierarchy of the **Event** portion of the SN could be partitioned into the eight ST groups shown in Fig. 1. Each semantic-type group is represented by an MST in the corresponding metaschema. An MST PHENOMENON OR PROCESS is defined to represent the ST group rooted at **Phenomenon or Process** in Fig. 1. The metaschema hierarchy derived from the partition in Fig. 1 was shown in Fig. 2.

Overall, a diagram of a metaschema serves as a good visualization mechanism supporting orientation to the SN and, in turn, the META, and helps in the navigation of the UMLS knowledge. In [12] we introduced various partial graphical views of groups of STs supported by the metaschema paradigm. These views can help in orientation of a user to the full scope of the SN's semantic relationships.

2.2. Top-down heuristic metaschema

An important assumption underlying the construction of the algorithmically generated metaschemas in previous research [12,15] is that the resulting subject areas of the SN are natural to a human. In order to validate this assumption, we conducted the following study [15]. A number of experts with reputation in the UMLS research area or related areas were selected. We sent each expert a diagram of the SN's IS-A hierarchy, i.e., the two trees rooted at **Event** and **Entity**.

The experts were asked to partition the SN starting at the roots (i.e., top-down). The design of the study follows the Aristotelian [30] paradigm, where species are specified according to genus and differentiate. Partitioning is done based on the extent of the difference between the child ST and its more general parent ST. Details of this study have appeared in previous publications [12,16] and are omitted.

The design of the metaschema utilizes the one-to-one correspondence between the ST groups underlying the MSTs, and their root STs. By selecting a set of STs that are "important and quite different" from their parents, a participating expert induces a partition of the SN, where each selected semantic type is a root of its group, implying a corresponding "expert metaschema."

While studying responses from our 11 UMLS experts, we found that individual participants' responses varied greatly both in the choice of STs marked as roots of groups and their numbers. For example, experts 1 and 2 chose 21 and 34 STs to name MSTs in their expert metaschemas, respectively. Table 1 shows the number of MSTs for each expert metaschema, corresponding to the number of STs marked by that expert with a minimum, maximum and average numbers of 12, 36 and about 26, respectively. The standard deviation is 10.23.

We were interested in quantifying the variability of the experts' responses. Towards this end, we computed the X-by-X agreement matrix, among X experts, to examine the agreement between any two experts in the same study group. In the agreement matrix, the number in row i and column j indicates how many MSTs expert i and expert j agree on. The agreement matrix of all 11 experts (Table 2) demonstrates the high variability of participant responses. For instance, participants 2 and 5 both marked 34 STs and agreed on 27 of them. The average inter-participant agreement is 16.76 (only about 63% of the average number of marked STs, 26.73), with a high of 30 and a low of 6. The large range shows the high variability of participant responses.

It was expected that some choices would be made by many participating experts. We were interested in metaschemas that represent a kind of aggregation of the experts' responses rather than in the expert metaschemas of the individuals. In particular, we constructed a sequence of cumulative metaschemas, each of which reflects a specific level of aggregation of the experts. Suppose we have X experts' responses. We define a threshold value N in the range $(1, X)$ to represent the level of aggregation. We construct the cumulative metaschema for a given N as follows. For each ST marked by at least N participating experts, an MST is defined and given the name of the ST. Then *meta-child-of*'s and *meta-relationships* are derived as described before.

In our study, we received responses from 11 experts ($X = 11$) and thus obtained 11 cumulative metaschemas by varying N over the range $(1, 11)$. For $N = 8$, for example, the same 16 STs were marked by at least 8 out of the 11 experts, and so the corresponding cumulative metaschema has 16 MSTs. Table 3 shows the number of semantic

Table 1
Number of MSTs each expert chose in the top-down study

Participant	1	2	3	4	5	6	7	8	9	10	11	Average
# MSTs (expert)	21	34	21	35	34	35	25	26	12	15	36	26.73

Table 2
Inter-participant agreement matrix; average = 16.76

	1	2	3	4	5	6	7	8	9	10	11
1		19	15	16	15	19	12	11	11	12	20
2			18	28	27	27	20	19	12	14	28
3				16	16	17	14	9	10	10	18
4					28	26	23	21	8	10	30
5						27	20	20	8	10	27
6							19	22	10	14	27
7								14	8	7	24
8									6	9	18
9										9	11
10											13

Table 3
Number of semantic types marked by at least N participants

Threshold (N)	1	2	3	4	5	6	7	8	9	10	11
# Marked MSTs	45	45	45	42	36	26	20	16	10	7	2

types marked for each N. Obviously, the larger the value of N, the smaller the common number of MSTs.

As we can see from Table 3, the number of MSTs varies from 2 (for $N = 11$) to 45 (for $N = 1$, 2 and 3). The corresponding metaschema for the $N = 11$ case contains only two MSTs ENTITY and EVENT, each spanning the whole corresponding tree of the SN. For the $N \leqslant 3$ cases, each non-leaf ST names an MST. The metaschema that emerges in those cases is effectively just the SN itself, without its leaves. No real grouping of related STs occurs. Obviously such extreme metaschemas are not interesting.

We denote the cumulative metaschema with the threshold value N representing a simple majority of the experts (i.e., $N = \lceil X/2 \rceil$) as the *consensus metaschema* [31]. For the top-down study, the consensus metaschema ($N = 6$) contains 26 MSTs. Its hierarchy is shown in Fig. 4. We call this metaschema the *top-down consensus metaschema*. Nevertheless, we were not satisfied by this semantic metaschema as reflecting experts' opinions. The variations among the various experts' metaschemas were too wide. For example, there were no STs that all experts chose, except for the roots, **Entity** and **Event** which were dictated by the top-down approach. We note that in the study, the number of MSTs varied greatly. Also, important MSTs such as MOLECULAR SEQUENCE(4), which appears in the lexical metaschema, or ANATOMICAL ABNORMALITY(3), ANIMAL(9), and PLANT(2) which appear in the cohesive metaschema, are missing from the top-down consensus metaschema, although we expected them.

One possible explanation for the large variations and the lack of expected MSTs is that the instructions we provided to the experts were deficient. The decision of choosing an ST to head a group is made without considering the group members (which are further down in the tree of SN) and thus were not scanned yet, in the top-down scanning. The decision is made solely based on comparing the current ST to its parent (P) and deciding whether it is too important and different from P to be in P's group. The exact instruction given was "while scanning, marked by star semantic types which you judge as important and quite different from their parent semantic types". No further elaboration was given.

In other words, the domain experts were asked to identify substantial or unusual shifts of granularity in the taxonomy of SN. The lower level of agreement seems to be caused by the need of the expert to make such decisions without full knowledge. By this we refer to the inherent problem of choosing the current ST to represent a group of descendant STs, without knowing who will be the members of the group since they were not scanned yet. Furthermore, although it is known that the group members are descendants of the current ST, it is not known which descendants, as some of them may be selected later for subsequent groups.

This problem led us to consider an alternative approach, where the SN is scanned bottom-up and a decision regarding a root of a group is made by an expert considering the whole group as well as the comparison to its parent. This was expected to enable experts to create a better metaschema based on semantic considerations, as their decisions would be based on more knowledge. In the next section, we present the bottom-up approach.

3. Methods

3.1. Design of bottom-up heuristic metaschema

3.1.1. The bottom-up study

In order to find a better way for validating algorithmically generated metaschemas, we conducted a bottom-up study with 13 participants. Instead of scanning the STs of SN from the roots to the leaves as in the top-down study, the participants in the bottom-up study were instructed to scan the two SN trees from the leaves up to the roots. The detailed instructions are as follows:

1. "A leaf (semantic type without children) is not chosen to head a group.
2. When processing the current semantic type, consider to what extent the descendant semantic types of its group are more specific than its parent semantic type. If it is much more specific, then choose the current semantic type to head its group by marking it with a star. That

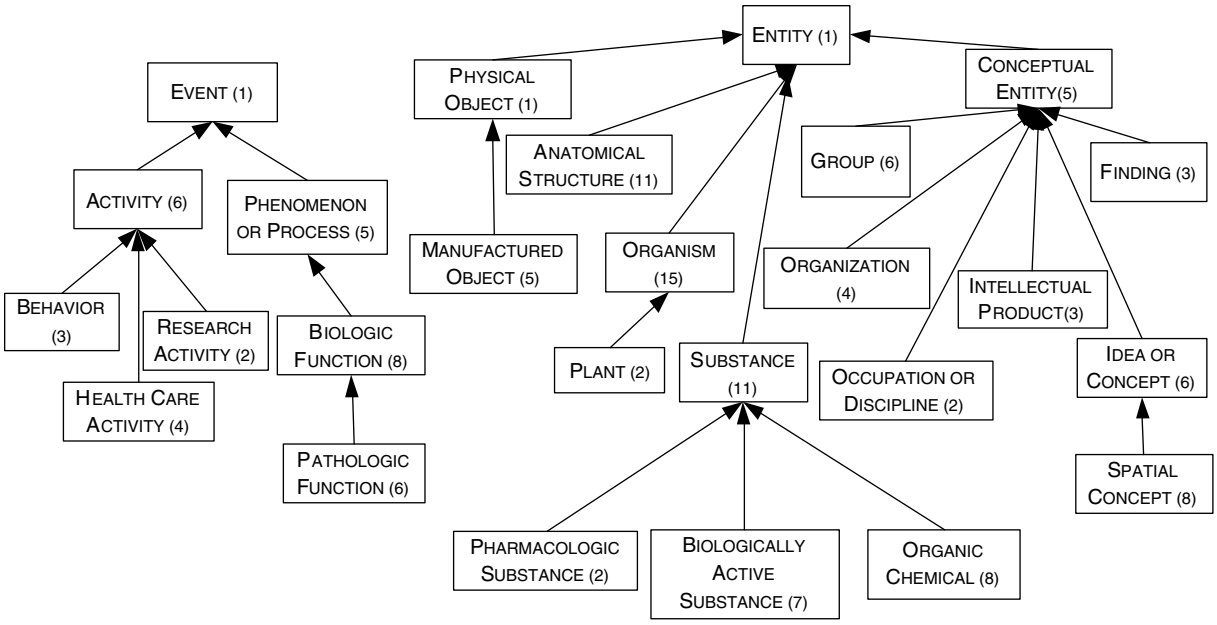


Fig. 4. Top-down consensus metaschema hierarchy.

means the parent will be in a separate group. Otherwise, the parent semantic type should be added to the group of the current semantic type.*Remark:* Although the marking is attached to the current semantic type, the decision is actually whether to include the parent in the same group.

3. The star marking of each participant will be used to define a metaschema, where each semantic type marked by a star names a metasemantic type. The metaschema will be compared with the results of other respondents and with our algorithmically derived metaschema''.

3.1.2. Reliability of the experts

We used reliability theory [32] to assess the variability of the experts in producing the bottom-up metaschema. When experts designate each ST as belonging or not belonging to a metaschema, their answers may reflect an idealized consensus opinion about which STs truly belong to it, or they may reflect error, noise, and differences of opinion. Cronbach's alpha [32] estimates the proportion of the total variability in the experts' answers that is due to true differences among STs (some do and some do not belong in the metaschema). The remainder of the variability (1 − Cronbach's alpha) represents the error, noise, and differences of opinion. Cronbach's alpha ranges from 0 to 1, where 0 represents pure noise and 1 represents perfect consensus among the experts. A value of .7 is often used as a target for reasonable reliability.

In general, by combining the answers of several experts, one obtains a better and more reliable result. One can report the average per rater reliability, which measures the degree to which the average expert tends to agree with other experts, or one can report the reliability of the combined result, which will always be higher than the per rater estimate. For example, a metaschema produced by one expert will not be as good as one produced by the combined opinion of several experts (a consensus metaschema). Cronbach's alpha estimates the reliability of the combined result, but the per rater reliability can be calculated easily from it [32].

We calculated Cronbach's alpha for the bottom-up metaschema, treating each of the 45 candidate STs as expert tasks. We then compared the result to that of the top-down metaschema. These results appear in Section 4.1.

3.2. Comparing metaschemas

In the comparison of two metaschemas, we consider not only the MST names, but also the underlying ST groups represented by the MSTs. To support the comparison, we present four definitions as follows.

Let M_A and M_B be two metaschemas of the SN.

Definition 1 (*Identical*)**.** An MST A in M_A is *identical* to an MST B in M_B if both MSTs have the same underlying ST group.

Since the ST group of an MST is connected and is part of the tree hierarchy of SN, this group is a tree. Since the root is used to name the MST of the group, both MSTs A and B share the same name.

Definition 2 (*Similar*)**.** An MST A in M_A is *similar* to an MST B in M_B if both MSTs have the same name and the same root.

Again, the names are the same.

This definition expresses that the names of two similar MSTs are equal. To better understand the differences between pairs of similar MSTs, we note that in some cases

the difference reflects various levels of granularity in the partition, rather than major disagreements between the metaschemas. An MST in one metaschema may be split into several MSTs in the other metaschema.

We now define "refinement" as follows. Let G_M(A) denote the ST group represented by the MST A in the metaschema M.

Definition 3 (*Refinement*)**.** Let A be an MST in metaschema M_A. If there exists a set of MSTs $\{B_1, B_2, \ldots, B_k\}$ ($k \geqslant 2$) in metaschema M_B such that A and B_1 (which is the root of $\{B_i\}$) are similar (that is, the STs A and B_1 are equal) and $G_{M_A}(\mathrm{A}) = \cup_{i=1}^{k} G_{M_B}(B_i)$, then the set $\{B_1, B_2, \ldots, B_k\}$ is called a *refinement* of A.

Definition 4 (*Refinable*)**.** Two similar MSTs A in metaschema M_A and B in the metaschema M_B are called refinable if either A has a refinement in M_B or B has a refinement in M_A.

Definition 5 (*Non-refinable*)**.** Two similar MSTs A in the metaschema M_A and B in the metaschema M_B, neither of which has a refinement in the other metaschema, are called non-refinable.

To illustrate these definitions, Fig. 5 demonstrates an abstract Semantic Network S of STs (Fig. 5a) and the two abstract metaschemas M_A (Fig. 5b) and M_B (Fig. 5c).

A black shadow for two MSTs with identical names in the two metaschemas M_A and M_B indicates identical MSTs. For example, the ST set for MST S1 is {S1} for both M_A and M_B. Both MSTs S2 and S3 for the metaschemas M_A and M_B are similar. But their characteristics differ. The occurrences of S2 in both metaschemas define a refinement. More precisely, $\{\mathrm{S2}(3), \mathrm{S5}(2)\}$ in M_A is a refinement of S2(5) in M_B, since the ST group of $\mathrm{S2}(5) = \{\mathbf{S2}, \mathbf{S4}, \mathbf{S6}, \mathbf{S5}, \mathbf{S9}\}$ is equal to the union $\mathrm{S2}(3) \cup \mathrm{S5}(2) = \{\mathbf{S2}, \mathbf{S4}, \mathbf{S6}\} \cup \{\mathbf{S5}, \mathbf{S9}\}$. The occurrences of **S2** in both metaschemas are refinable. The occurrences of **S3** in both metaschemas are non-refinable.

We compare two metaschemas using the above three terms to measure the similarity between their ST coverages. To capture cases of either identical MSTs or MSTs which reflect only granularity differences between two metaschemas, we introduce another term, *correspondable* MSTs.

Definition 6 (*Correspondable*)**.** An MST A in M_A is *correspondable* to an MST B in M_B if A and B are either identical or refinable.

Definition 7 (*Corresponding MST groups*)**.** Two groups of MSTs in two metaschemas M_A and M_B, respectively, are *corresponding MST groups* if either both groups are singletons of identical MSTs or one group is a singleton and the other group is a refinement of the MST of the singleton.

For example, in Fig. 5 there are two pairs of corresponding MST groups, shown by broken lines circumscribing them. They are the identical S1(1) in M_A and S1(1) in M_B and the groups $\mathrm{S2}(3) \cup \mathrm{S5}(2)$ in M_A and S2(5) in M_B of which the first is a refinement of the second.

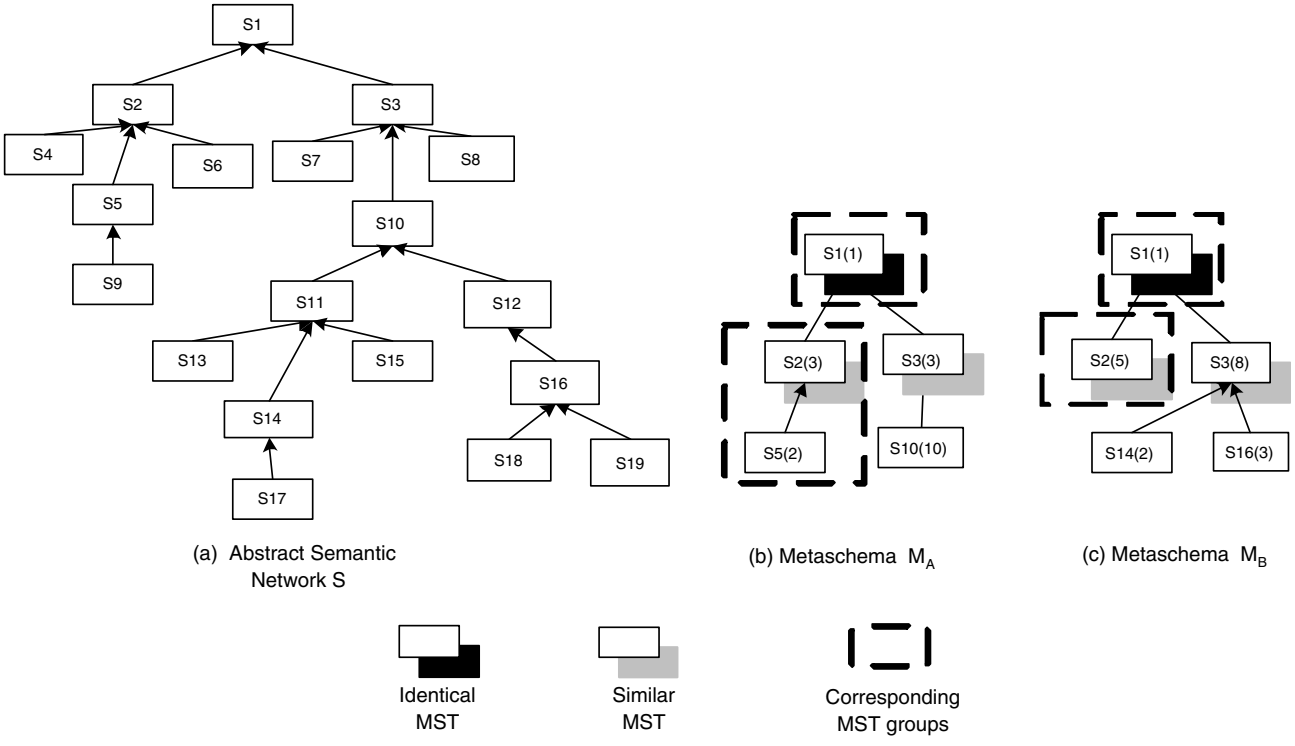


Fig. 5. The abstract Semantic Network and input metaschemas.

3.3. Structural properties of metaschemas

We will now list several structural metrics for characterizing a metaschema M.

(1) Cardinality C: The number of MSTs in a metaschema.
(2) Complexity: The ratio of the number of relationships (both hierarchical and semantic relationships) to the cardinality.For convenience we will refer to the number of STs represented by an MST M_i, $i = 1, \ldots, C$ as the weight $W(M_i)$ of the MST.
(3) Maximum weight $MAXW = \max_{1 \leqslant i \leqslant C} W(M_i)$.
(4) Minimum weight $MINW = \min_{1 \leqslant i \leqslant C} W(M_i)$.
(5) Weight spread $WS = MAXW - MINW$.
(6) Average weight $AVGW = \frac{1}{C}\sum_{i=1}^{C} W(M_i)$.
(7) Standard deviation of the weights of the MSTs. $\sigma = \sqrt{\frac{\sum_{i=1}^{C}(W(M_i)-AVGW)^2}{C}}$.
Note that the standard deviation's contribution of an MST M_i, defined as $SDC(M_i) = (W(M_i) - AVGW)^2$, is evaluated when comparing the standard deviations, since $\sigma = \sqrt{\frac{\sum_{i=1}^{C} SDC(M_i)}{C}}$.
(8) Coverage: Percentage of SN semantic relationships covered by the meta-relationships of the metaschema. This measure is based on [15,16] and will be used sparingly in this paper.

From the structural point of view, in an ideal partition of n elements into k groups, each group will have an almost equal weight (WS is at most 1). Such a partition is called a *uniform partition*. However, if the elements are nodes of a tree and the partition is into connected subtrees, then due to the structure of the tree, a uniform partition is not always possible. Thus, as an approximation to a uniform partition, we are looking for a partition with a minimum weight spread. Other alternatives are a partition with a minimum heaviest weight (MIN–MAX partition) or a partition with a maximum lightest weight (MAX–MIN partition). For algorithms to construct a MAX–MIN partition and a MIN–MAX partition of a weighted tree, see [33,34]. Beyond the two extreme measures, $MAXW$ and $MINW$, of the partition, we would like all its weights to be as close as possible to the average weight. For this purpose, we want the standard deviation of the weights to be as small as possible.

Furthermore, the partition underlying a metaschema will probably not be uniform due to its need to capture different subject areas correctly. This is a much more important consideration than the equal size of the MST groups. Nevertheless, there are cases where one can choose between two options regarding the grouping, for which there is no clear-cut semantic reason to decide between them. In such a case, we should follow the structural criteria and prefer the option which tends to equalize the weights of the groups, avoiding groups which are too large or too small.

3.4. Consolidation

We have developed an algorithm which takes two given metaschemas $M_A = \{A_1, A_2, \ldots, A_m\}$ and $M_B = \{B_1, B_2, \ldots, B_n\}$ for an abstract Semantic Network S as input, and generates an output M_C, a consolidated metaschema.

When constructing the consolidated metaschema, the algorithm attempts to minimize the $MAXW$ and the weights' standard deviation, while maximizing the $MINW$ for this metaschema. In doing this, the algorithm tries to improve the structural properties of the consolidated metaschema by choosing MSTs of the given metaschemas accordingly.

In the process of the algorithm, we will construct a sequence of auxiliary Semantic Networks and auxiliary metaschemas derived from the original Semantic Network S and metaschemas M_A and M_B. In the description of the algorithm, we will use the previously defined terms: identical MSTs, similar MSTs, refinable MSTs, correspondable MSTs, non-refinable MSTs and corresponding MST groups. We will need a few more definitions.

Definition 8 (*Auxiliary induced metaschema*)**.** Given a metaschema M_A defined for a Semantic Network S, an *auxiliary induced metaschema* M'_A is obtained from M_A by deleting some selected MSTs of M_A or by combining some groups of MSTs of M_A such that all the *child-of* in the original metaschema M_A among MSTs of M'_A exist in M'_A.

Definition 9 (*Expanded Semantic Network*)**.** Let M'_A be an auxiliary induced metaschema of the metaschema M_A defined for Semantic Network S. The expanded Semantic Network S′ of the metaschema M'_A contains all the STs of all MSTs of M'_A and all IS-A relationships in the original Semantic Network S among the STs of S′.

We note that S may consist of several connected components (as for the UMLS SN). In such a case the algorithm works independently on each component. The algorithm MAIN-CONSOLIDATE (M_A, M_B, S, M_C) takes two stages. It first invokes its core procedure R-CONSOLIDATE (M_A, M_B, S, M_C) to obtain an initial consolidated metaschema M_C, and then further modifies M_C to complete the consolidation. We will now present a high-level description of the algorithm MAIN-CONSOLIDATE (M_A, M_B, S, M_C), followed by a step-by-step description.

The procedure R-CONSOLIDATE (M_A, M_B, S, M_C) is a recursive procedure to create a consolidated metaschema M_C from two given metaschemas M_A and M_B of a Semantic Network S. The procedure starts by selecting all the identical MSTs of M_A and M_B for M_C. It continues by selecting from each pair of corresponding MST groups of M_A and M_B, an MST or a group of MSTs, which minimizes the standard deviation contributions, to be added to M_C. The identical MSTs and corresponding MST

groups of both input metaschemas are deleted from M_A and M_B in such a way that auxiliary induced metaschemas M'_A and M'_B are generated. Next we create the expanded Semantic Network S′ of M'_A (and of M'_B, which is identical).

At this stage, all root MSTs of M'_A and M'_B are non-refinable. Let A_i and B_j be two similar root MSTs of M'_A and M'_B, respectively. We add to M_C the one of A_i and B_j which minimizes the standard deviation contribution. Without loss of generality, assume that MST A_i of M'_A was selected for adding to M_C. We derive the induced auxiliary metaschema M''_A by removing MST A_i from M'_A. Next, we obtain the expanded Semantic Network S″ from M''_A. If M''_A (and S″) are empty, we return M_C and the algorithm is finished.

If M''_A is not empty, we face a difficulty. We cannot obtain the auxiliary metaschema M''_B from M'_B by deleting A_i, since A_i is not an MST in M'_B, neither does it have a refinement in M'_B, since the MST A_i of M'_A is non-refinable. As a result, we will obtain M''_B in an indirect way, following the derivation of several auxiliary induced metaschemas of M'_B and their expanded Semantic Networks.

First we obtain the auxiliary induced metaschema M^*_B from M'_B which will include all the MSTs of M'_B for which all their STs are in S″. The expanded Semantic Network S* of M^*_B is derived. The Semantic Network D is derived by deleting from S′ all the STs of S*. Next, we derive an auxiliary induced metaschema SD for the Semantic Network D consisting of the connected components of D. Each component is represented by one MST, named after the root. Finally, we derive the desired auxiliary induced metaschema M''_B by combining the metaschemas M^*_B and SD. At this stage, we are ready for a recursive call of the procedure R-CONSOLIDATE (M''_A, M''_B, S″, M_C), to update the metaschema M_C. After receiving the updated M_C, we return it as a partial result. If M_C contains an MST of one ST, with at most one child, it is added to its parent MST.

We will now describe this algorithm as a series of separate steps. To keep track of this fairly complicated process, we provide in Fig. 6 a diagram, which reflects the process described in R-CONSOLIDATE and the reader may use it as a road map. The procedure's steps are labeled by numbers. By necessity some of the numbers occur twice in the diagram, because they describe operations with two inputs or because the described operation may occur for either one of the two input metaschemas. Following the step-by-step algorithm description, there is an example. The reader is advised to review the example in parallel to reading the algorithm.

Procedure R-CONSOLIDATE (M_A, M_B, S, M_C)

Step 1: All MSTs that are identical (as defined above) in both input metaschemas M_A, M_B are included in the output metaschema M_C.

Step 2: When given an ST in one input metaschema and its refinement in the other input metaschema, for example, the set $\{B_1, B_2, \ldots, B_k\}$ with a *refinement* of A_j, then if $\sum_{i=1}^{k} SDC(B_i) < SDC(A_j)$, include the set $\{B_1, B_2, \ldots, B_k\}$ in the output metaschema M_C. Otherwise, include A_j in M_C.The same rule applies to an MST B_l with a *refinement* $\{A_1, A_2, \ldots, A_m\}$.
We note that in case of different cardinalities for M_A and M_B we use the average of the two cardinalities for calculating the contribution to the standard deviation.

Step 3: We construct two auxiliary induced metaschemas M'_A (M'_B) from M_A (M_B) by removing from M_A (M_B) all corresponding MST groups (identified in the two previous steps).

Step 4: We construct an expanded Semantic Network S′ of M'_A and the given Semantic Network S.

Step 5: Choosing from non-refinable similar root MSTs A_i and B_j (in M'_A and M'_B, respectively), (i) if $SDC(A_i) < SDC(B_j)$, include A_i in the consolidated metaschema M_C. Otherwise, (ii) include B_j in M_C. If there are several roots, such a choice is made for each root.

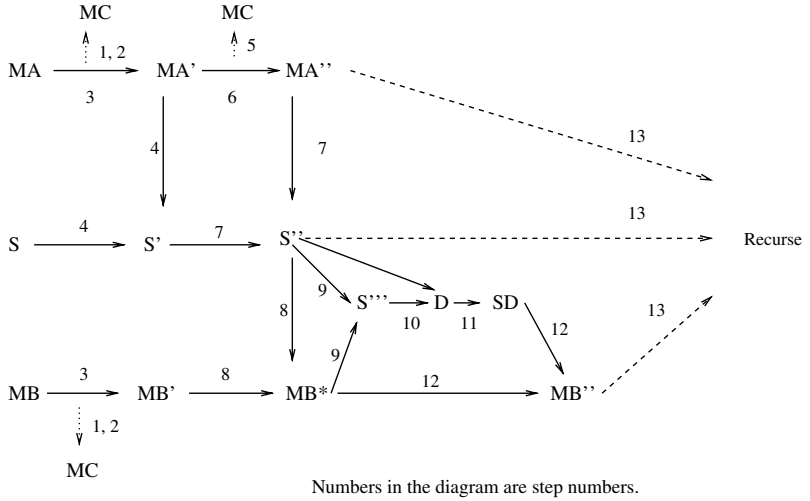


Fig. 6. The flow chart of consolidation numbered steps.

Step 6: Assuming without loss of generality that A_i of M'_A was selected (case (i)), we construct an auxiliary induced metaschema M''_A from M'_A by removing from M'_A the root MST A_i. If M''_A is empty then return. Note that, if B_j was selected for case (ii), the respective role reversal of M'_A and M'_B would follow.

Step 7: We construct an expanded Semantic Network S″ of M''_A and the Semantic Network S′.

Step 8: We construct an auxiliary induced metaschema M^*_B from M'_B as follows. Only those MSTs (from M'_B) which have all their semantic types in S″ are included in M^*_B.

Step 9: We construct an expanded Semantic Network S* of M^*_B and the Semantic Network S″.

Step 10: We construct the difference D of S″ and S* as follows. D contains all the STs of S″ which are not in S*, i.e., we use traditional set difference.

Step 11: We construct an auxiliary induced metaschema SD, which consists of the maximally connected components of D, with each component corresponding to one MST, named after its root.

Step 12: We construct the auxiliary induced metaschema M''_B as follows. We combine SD with M^*_B using appropriate *meta-child-of* relationships from the MSTs of M^*_B up to the MSTs of SD.

Step 13: If the metaschemas M''_A and M''_B consist of one connected component then recursively call R-CONSOLIDATE (M''_A, M''_B, S'', M_C). Otherwise, recursively call R-CONSOLIDATE for every pair of connected components (M''_{A_i}, M''_{B_j}), such that M''_{A_i} and M''_{B_j} have similar MST roots, A_i and B_j, with their corresponding expanded Semantic Network.
Return the partially consolidated metaschema M_C.

END Procedure R-CONSOLIDATE

Algorithm MAIN-CONSOLIDATE (M_A, M_B, S, M_C)

Stage 1: Invoke R-CONSOLIDATE (M_A, M_B, S, M_C).

Stage 2: (Modification): Each MST in M_C with only one ST (such MSTs are called singletons) is combined with its parent MST whenever this child MST is a leaf or has a single child in M_C.
Return the completely consolidated metaschema M_C.

END Algorithm MAIN-CONSOLIDATE

3.4.1. Example

We demonstrate the MAIN-CONSOLIDATE (M_A, M_B, S, M_C) algorithm for M_A, M_B and S given in Fig. 5.

The algorithm first invokes procedure R-CONSOLIDATE, passing M_A, M_B, S and M_C as arguments. The following Steps 1–13 show the process of R-CONSOLIDATE.

Step 1: S1 is an identical MST in M_A and M_B and is included in M_C (Fig. 7a–c).
There are 19 STs in the abstract Semantic Network S, and 5 MSTs in both M_A and M_B. Thus, the average MST's weight for both metaschemas is 3.8, which is used for computing the SDC.

Step 2: S2(5) in M_B (Fig. 7b) contributes 1.44 to the standard deviation, while its refinement {S2(3), S5(2)} in M_A (Fig. 7a) contributes $0.64 + 3.24 = 3.88$. In this case, S2(5) of M_B is chosen for M_C (Fig. 7c).

Step 3: S1(1), S2(3) and S5(2) are removed from M_A to yield M'_A. S1(1) and S2(5) are removed from M_B to yield M'_B (Fig. 7d and e).

Step 4: We construct S′ by expanding of S3(3) and S10(10) of M'_A (Fig. 7f).

Step 5: For our example, S3(3) in M'_A (Fig. 7d), which contributes 0.64 to the standard deviation, is chosen, rather than S3(8), in M'_B (Fig. 7e), which contributes 17.64.

Step 6: The MST S3(3) of M'_A is deleted as is indicated in Fig. 8a by dotted lines, to obtain M''_A as shown in Fig. 8d.

Step 7: S″ is generated by reexpanding M''_A, which has one MST S10(10) (see Fig. 8e).

Step 8: For M'_B of Fig. 9a and S″ of Fig. 9b, $M^*_B = \{S14(2), S16(3)\}$, see Fig. 9c. S3(8) is not included in M^*_B, because S3 is not in S″.

Step 9: We expand M^*_B to get S*. For the M^*_B of Fig. 9c, S* is shown in Fig. 9d.

Step 10: D is constructed by removing S14(2) and S16(3) from S″ (Fig. 9e).

Step 11: As there is only one component, the induced auxiliary metaschema SD consists of S10(5) only (Fig. 9f).

Step 12: S10(5) is combined with S14(2) and S16(3) into M''_B (Fig. 9g). We note that the MST S10(5) was not an MST in the original M_B metaschema.What have we achieved now? M''_B is a metaschema for S″. M''_A has been a metaschema of S″ all along. Most importantly, neither M''_A nor M''_B represent any semantic types which are represented by the MSTs in M_C. Thus, we have extended the output metaschema, while the two input metaschemas have been shrunk correctly, and they correctly summarize their corresponding "shrunk" Semantic Network.

Step 13: The MST group S10(5), S14(2), S16(3) of M''_B, in Fig. 9g, is the refinement of S10(10) of M''_A in Fig. 8d. In this case the refinement in M''_B has a lower standard deviation contribution than S10(10) and is chosen for the consolidated metaschema. After the deletion of the corresponding

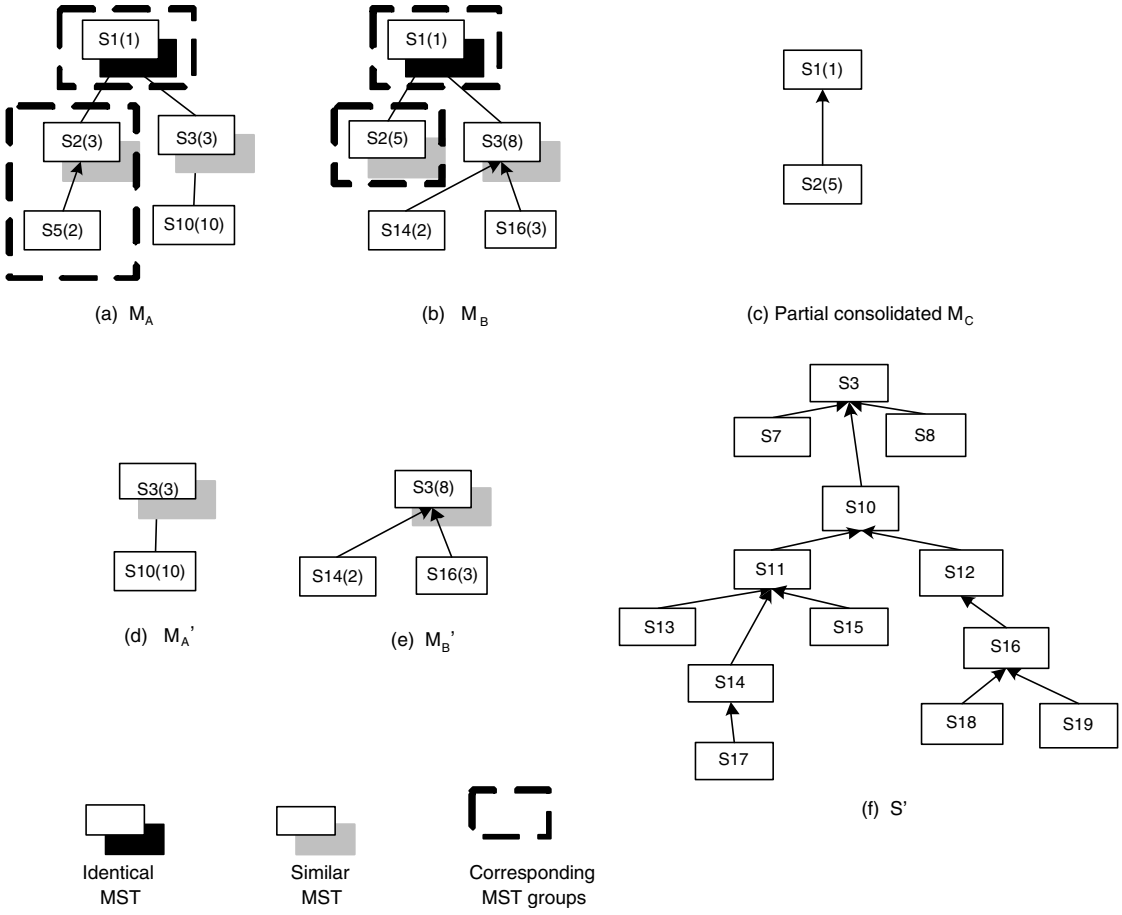


Fig. 7. Consolidating and deleting corresponding MSTs.

MSTs, the resulting metaschemas are empty and the procedure returns.

At this point, the M_C is a partially consolidated metaschema and the first stage of the MAIN-CONSOLIDATE algorithm is done. At the second stage, we see S1(1) is a singleton MST, but it has two children, thus it is legitimate and no modification occurs. The consolidated metaschema M_C for the given metaschemas M_A and M_B for the abstract Semantic Network S is shown in Fig. 10.

4. Results

4.1. Bottom-up heuristic metaschema

In the bottom-up study, we received responses from 13 experts. Individual participants' responses varied both in the choice of STs marked and their numbers. For example, experts 1 and 3 chose 28 and 17 STs, respectively, to name MSTs in their expert metaschemas. Table 4 shows the number of MSTs for each expert metaschema, corresponding to the number of STs marked by that expert. The average number of MSTs marked is 23, with minimum and maximum numbers of 16 and 30, respectively. The standard deviation is 4.56.

Each expert's response can be used to construct one expert's metaschema. We obtained 13 cumulative metaschemas, from the 13 experts' ($X = 13$) metaschemas, by varying N over the range (1,13). In the Nth cumulative metaschema, $N = 1, \ldots, 13$, each MST was chosen by at least N experts. For $N = 8$, for example, there were 16 STs marked by at least 8 out of the 13 experts, and so the corresponding cumulative metaschema has 16 MSTs. Table 5 shows the number of STs marked for each N.

As we can see from the table, the number of MSTs varies from 3 (for $N = 13$) to 41 (for $N = 1$ and 2). Obviously such extreme metaschemas are not interesting. The consensus metaschema ($N = 7$) contains 25 MSTs. Its hierarchy is shown in Fig. 11.

We observed that in the bottom-up study too, individual participants' responses varied greatly, both in the choice of STs marked and their numbers. To substantiate this, we constructed the agreement matrix of all 13 experts (Table 6) that demonstrates the agreement as well as the high variability of participant responses. For instance, participants 1 and 4 marked 28 and 25 STs, respectively, and agreed on only 16 of them. The average inter-participant agreement is 14.41 (only 67% of the average number of 23 marked STs), with a high of 25 and a low of 6.

Cronbach's alpha for the consensus bottom-up metaschema was .79. This implies that the consensus metasche-

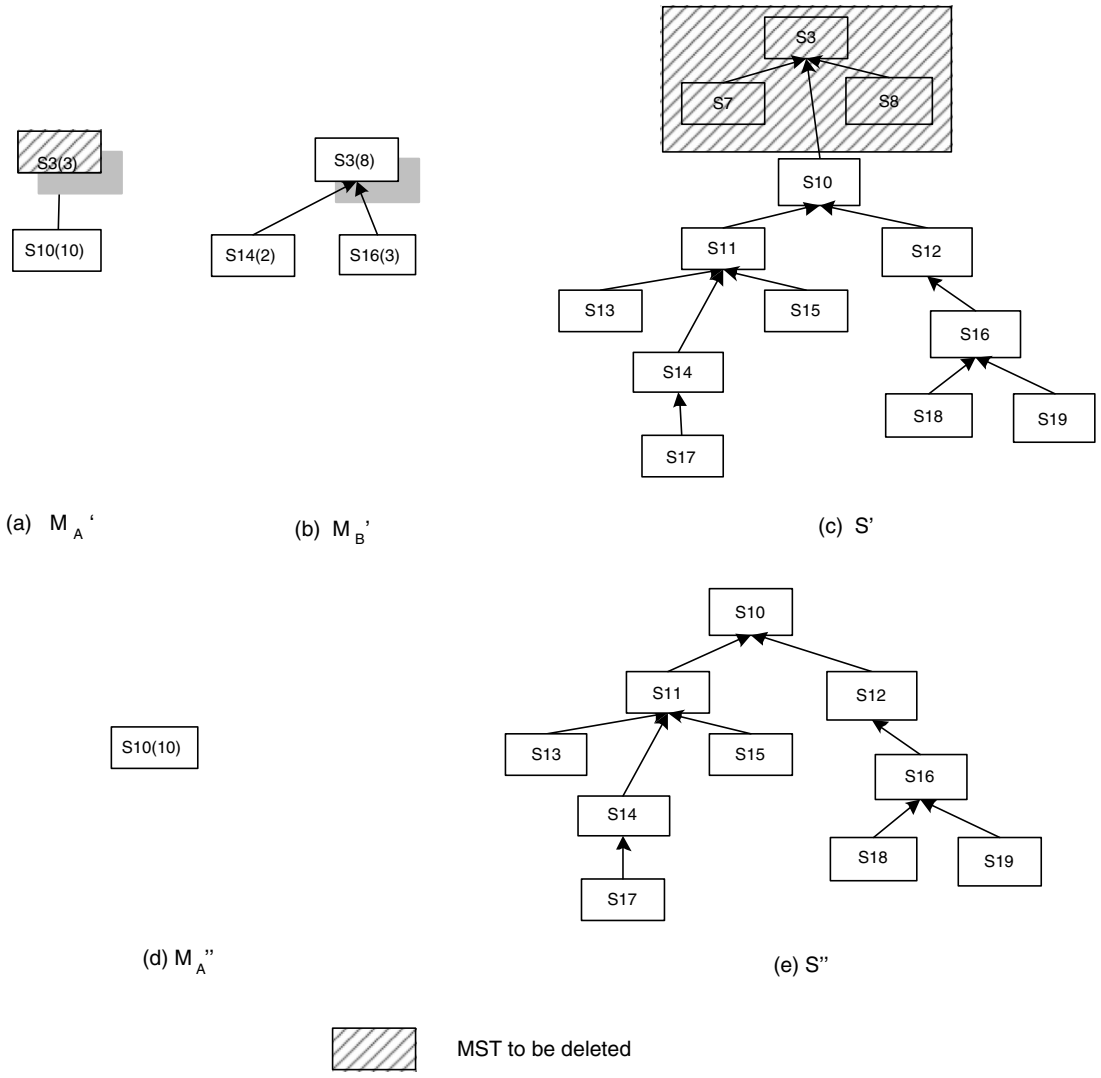


Fig. 8. Constructing M''_A and S″.

ma is sufficiently reliable (greater than .7; see Section 3.1.2). For the consensus top-down metaschema, Cronbach's alpha was .62, which is lower than the threshold, but still reasonable.

By looking at the per rater reliability, one can correct for the fact that the bottom-up metaschema had more experts than the top-down one. The per rater reliability for the bottom-up metaschema was .23, and for the top-down metaschema it was .13. The difference was borderline ($p = .053$). These results imply that a metaschema produced by a single expert by either method is insufficiently reliable (i.e., both are well below .7) and that the bottom-up approach is probably more reliable than the top-down approach, although the difference did not quite achieve statistical significance.

Another way to understand the results is to ask how many experts' answers would need to be combined to achieve the target reliability of .7. The bottom-up approach would require 8 experts on average, whereas the top-down approach would require 16.

4.2. Results of metaschema comparison

To facilitate the comparison between the consensus metaschemas obtained from the two studies, we show both their hierarchies in Fig. 12. MSTs identical in both metaschemas are indicated by black shadows. Similar MSTs are denoted by gray shadow.

There are 12 MSTs identical for the two metaschemas. For example PATHOLOGICAL FUNCTION(6) is an MST in both metaschemas, representing the same underlying ST group. Table 7 lists all the identical MSTs and their sizes. Hence, both metaschemas agree that these 12 MSTs represent important subject areas in the SN. Altogether, they cover 47 STs (i.e., 34.8% of the SN).

There are seven similar MSTs. For example, BIOLOGICAL FUNCTION(8) in the top-down study metaschema is similar to BIOLOGICAL FUNCTION(1) in the bottom-up metaschema. Table 8 shows these similar MSTs along with their sizes in each of the two metaschemas. In the top-down study metaschema, these seven MSTs cover 60

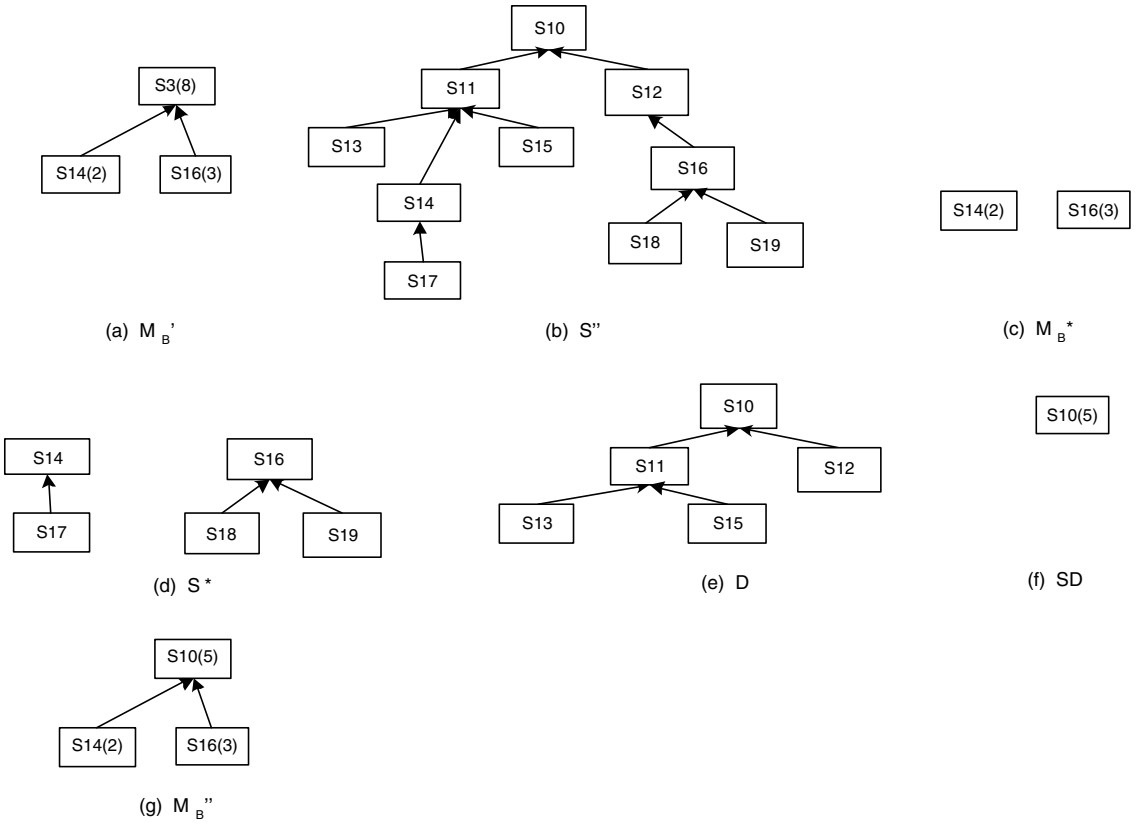


Fig. 9. Constructing M''_B.

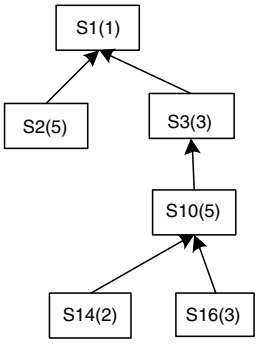


Fig. 10. Consolidated metaschema.

STs, which is about 44% of the SN. In the bottom-up study metaschema, these seven MSTs cover 38 STs, which is about 28%.

To better understand the nature of the similarity represented in Table 8, we will explore refinements in both directions. As a refinement of the top-down consensus metaschema, consider the MST ORGANISM(15) in the top-down consensus metaschema. This MST is split into two separate MSTs, ORGANISM(6), and ANIMAL(9), in the bottom-up consensus metaschema. In other words, {ORGANISM(6), ANIMAL(9)} in the bottom-up consensus metaschema is a refinement of ORGANISM(15) in the top-down consensus metaschema. The refinement cases cover 42 STs in both metaschemas. Table 9 lists the cases of refinement of the top-down consensus metaschema.

Considering refinements in the other direction of the bottom-up consensus metaschema, there is one case. {ACTIVITY(6), RESEARCH ACTIVITY(2)} in the top-down study metaschema is the refinement of ACTIVITY(8) in the bottom-up study metaschema.

Table 4
Number of MSTs each expert chose in the bottom-up study

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	Average
# MSTs	28	25	17	25	22	25	19	16	19	24	20	29	30	23

Table 5
Number of MSTs chosen by at least N participants

Threshold (N)	1	2	3	4	5	6	7	8	9	10	11	12	13
# MSTs	41	41	36	35	33	32	25	22	14	8	6	4	3

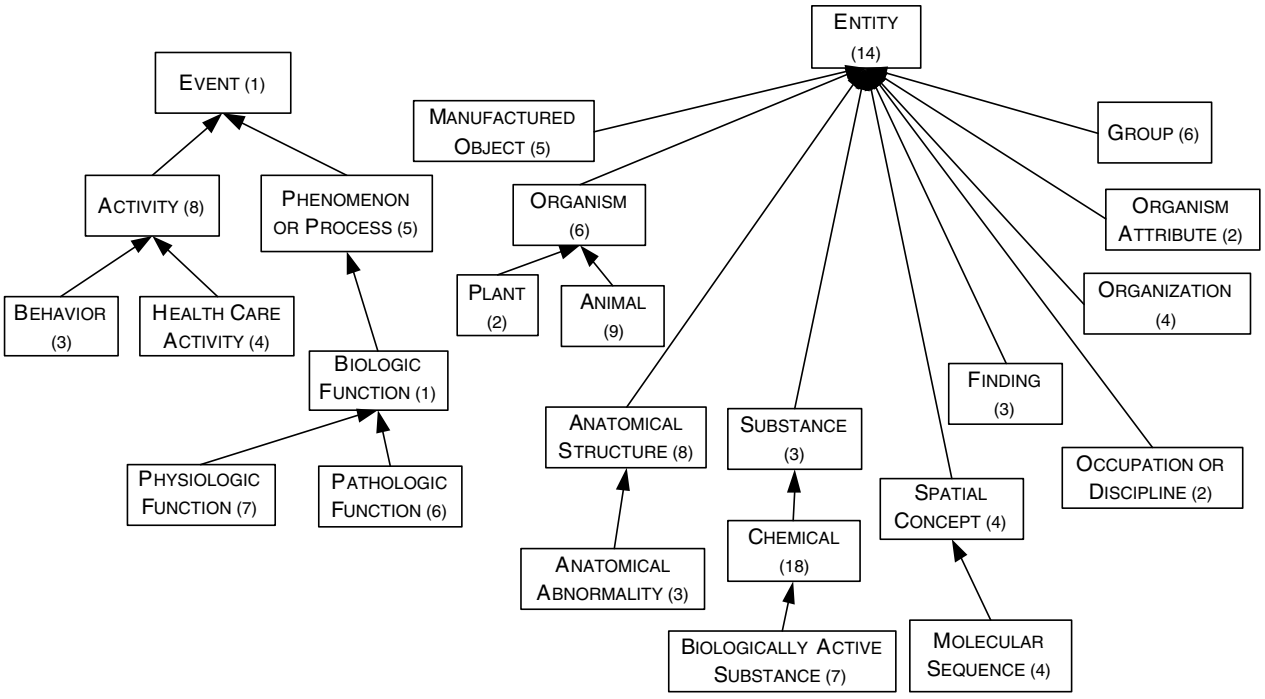


Fig. 11. Bottom-up consensus metaschema hierarchy.

Table 6
Inter-participant agreement matrix; average = 14.41

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		20	12	16	19	16	15	13	13	18	25	20	23
2			13	14	17	15	16	12	11	17	17	18	24
3				8	13	13	6	9	10	11	12	15	16
4					12	14	12	8	14	18	14	19	16
5						11	13	9	7	17	15	19	22
6							14	12	16	16	13	19	17
7								7	9	14	10	16	16
8									11	9	9	9	14
9										14	11	20	13
10											15	20	26
11												16	19
12													22

4.3. Results of structural evaluation

In Table 10, we show the values of the eight structural measures for both the consensus metaschemas. For example, the top-down consensus metaschema has the cardinality 26, while the bottom-up consensus metaschema has the cardinality 25. Table 11 shows the distribution of the weights for both metaschemas. For example, both metaschemas contain three MSTs of weight six. Here, weight six means there are six STs in the group represented by the MST.

4.4. Consolidated metaschema

Following the MAIN-CONSOLIDATE algorithm (see Section 3.4), we may derive a consolidated metaschema of the two consensus metaschemas as follows. Steps 1–13 show the process of the R-CONSOLIDATE procedure.

4.4.1. Step 1

All identical 12 MSTs in both metaschemas (marked with black shadows in Fig. 12) are included in the consolidated metaschema.

4.4.2. Step 2

We selected the MST or its refinement whichever minimizes the SDC. The two consensus metaschemas have different cardinalities, 25 and 26. Thus, in calculating the contribution to the standard deviation, the average 25.5 of the two cardinalities is used. For example, on one hand, BIOLOGICAL FUNCTION(8) in the top-down consensus metaschema contributes to the standard deviation 7.3, while the refinement {BIOLOGICAL FUNCTION(1), PHYSIOLOGICAL

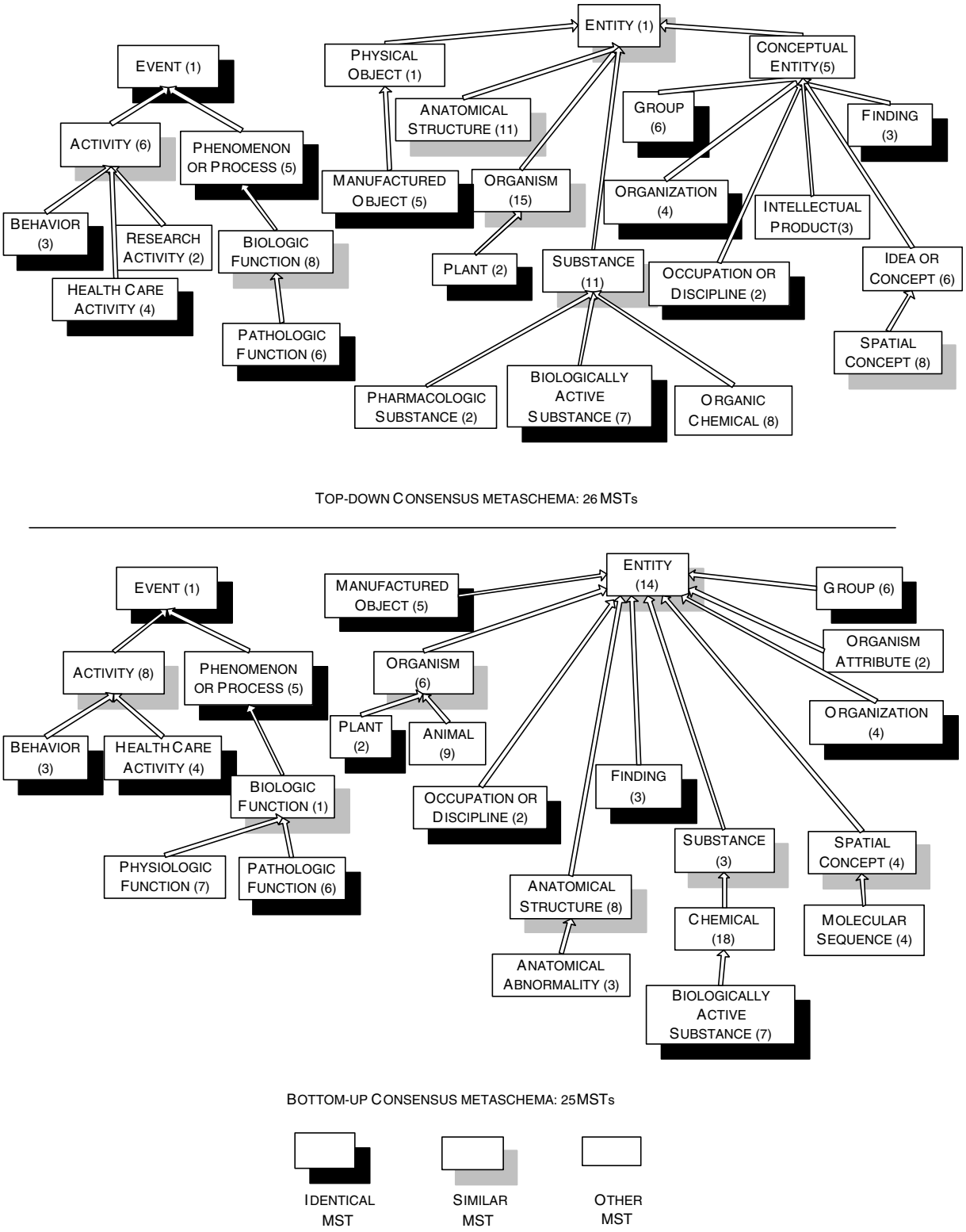


Fig. 12. Comparison of consensus metaschemas.

FUNCTION(7)} in the bottom-up consensus metaschema contributes 21.4. In this case, BIOLOGICAL FUNCTION (8) is selected for the consolidated metaschema. On the other hand, ORGANISM(15) in the top-down consensus metaschema contributes 94.1, but its refinement in the bottom-up metaschema {ORGANISM(6), ANIMAL(9)} only contributes 14.2. This refinement is selected for the consolidated metaschema. The corresponding MSTs are marked either by black shadows for identical MSTs or are marked with dashed borders for similar MSTs with refinements (see

Table 7
Identical MSTs in both metaschemas

MST	Size
BEHAVIOR	3
BIOLOGICALLY ACTIVE SUBSTANCE	7
EVENT	1
FINDING	3
GROUP	6
HEALTH CARE ACTIVITY	4
MANUFACTURED OBJECT	5
OCCUPATION OR DISPLIN	2
ORGANIZATION	4
PATHOLOGIC FUNCTION	6
PHENOMENON OR PROCESS	5
PLANT	2
Total: 12	47

Table 8
Similar MSTs in both metaschemas

MST	Weight in top-down study metaschema	Weight in bottom-up study metaschema
ANATOMICAL STRUCTURE	11	8
ACTIVITY	6	8
BIOLOGICAL FUNCTION	8	1
ENTITY	1	8
ORGANISM	15	6
SPATIAL CONCEPT	8	4
SUBSTANCE	11	3
Total: 7	60	38

Table 9
Refinements in bottom-up consensus metaschema

MST in top-down metaschema	Refinement in the bottom-up metaschema
ANATOMICAL STRUCTURE(11)	{ANATOMICAL STRUCTURE(8), ANATOMICAL ABNORMALITY(3)}
BIOLOGICAL FUNCTION(8)	{BIOLOGICAL FUNCTION(1), PHYSIOLOGICAL FUNCTION(7)}
ORGANISM(15)	{ORGANISM(6), ANIMAL(9)}
SPATIAL CONCEPT(8)	{SPATIAL CONCEPT(4), MOLECULAR SEQUENCE(4)}
Total: 4	42

Table 10
Values for the structural measures for the two consensus metaschemas

Measures	Top-down	Bottom-up
Cardinality	26	25
Complexity	4.3	4.72
Maximum weight	15	18
Minimum weight	1	1
Weight spread	14	17
Average weight	5.19	5.40
Standard deviation	3.49	3.93
Coverage (%)	70.6	75.93

Fig. 13a and b). Fig. 14 shows the partial consolidated metaschema after selecting corresponding MSTs, where the MST corresponding groups selected for the consolidated metaschema to minimize the SDC, are circumscribed by broken lines.

4.4.3. Step 3

We construct two auxiliary metaschemas M'_A and M'_B by deleting from M_A and M_B the identical MSTs and the corresponding MST groups, as shown in Fig. 15a and b.

4.4.4. Step 4

An expanded Semantic Network SN′ is derived using all STs summarized by all MSTs of M'_A together with the IS-A relationships directed to them in SN. Fig. 15c shows SN′, which consists of two subtrees, one rooted at **Substance** and the other at **Entity**. Each of the two auxiliary metaschemas M'_A and M'_B is a metaschema of SN′. The consolidation of each of the subtrees is now described independently. We note that at this stage the **Event** portion of the consolidated metaschema is fully determined.

4.4.5. Step 5 for *Substance*

For the subtree of SN′ rooted at **Substance** (Fig. 15c), based on the SDC, SUBSTANCE(3) in the bottom-up metaschema M'_B is chosen, rather than SUBSTANCE(11) in the top-down metaschema M'_A.

4.4.6. Step 6 for *Chemical*

Since SUBSTANCE(3) in the bottom-up metaschema is selected for the consolidated metaschema, an auxiliary induced metaschema M''_A is obtained, which contains only one MST, CHEMICAL(18).

4.4.7. Step 7 for *Chemical*

SN″ is the expanded Semantic Network for M''_A. For brevity, we will omit figures for some steps.

4.4.8. Step 8 for *Substance*

Since SUBSTANCE(3) is not an MST of the top-down metaschema, M^*_B needs to be constructed from M'_B in order to obtain M''_B. M^*_B contains the MSTs from M'_B which have all their STs in SN″, in this case, PHARMACOLOGIC SUBSTANCE(2) and ORGANIC CHEMICAL(8).

4.4.9. Step 9 for *Pharmacologic Substance and Organic Chemical*

We construct an expanded Semantic Network SN* from the metaschema M^*_B.

4.4.10. Step 10 for *Chemical*

The difference D of SN″ and SN* is constructed.

4.4.11. Step 11 for *Chemical*

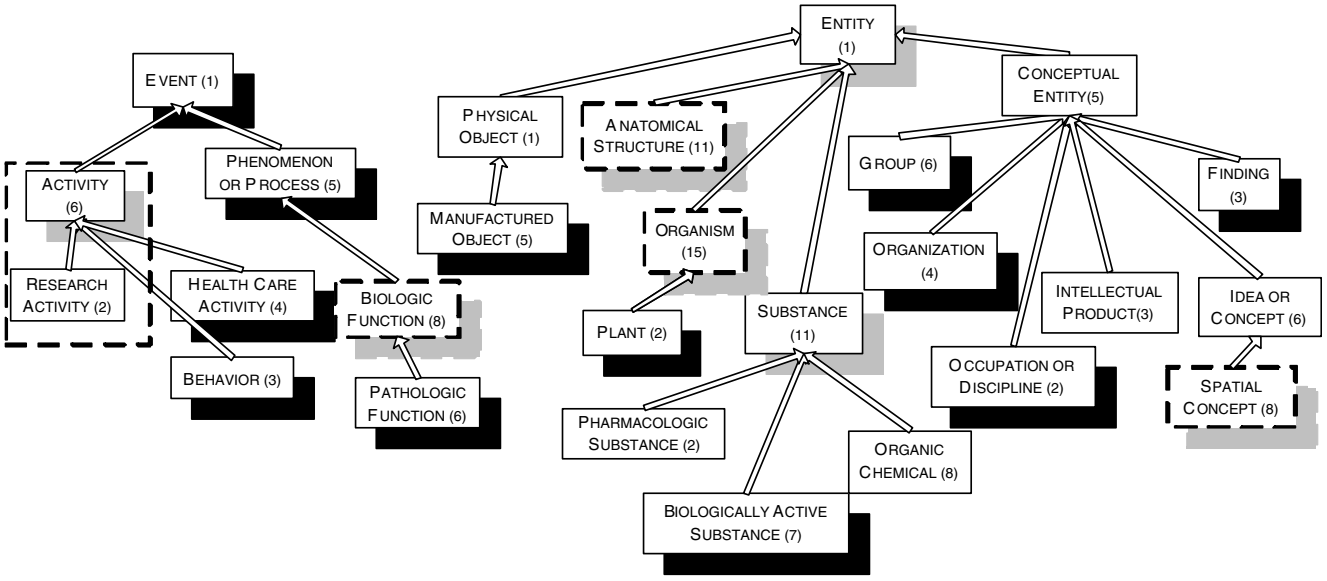
The SD now consists of CHEMICAL(8) obtained from the grouping of the STs which are contained in D.

4.4.12. Step 12 for *Chemical*

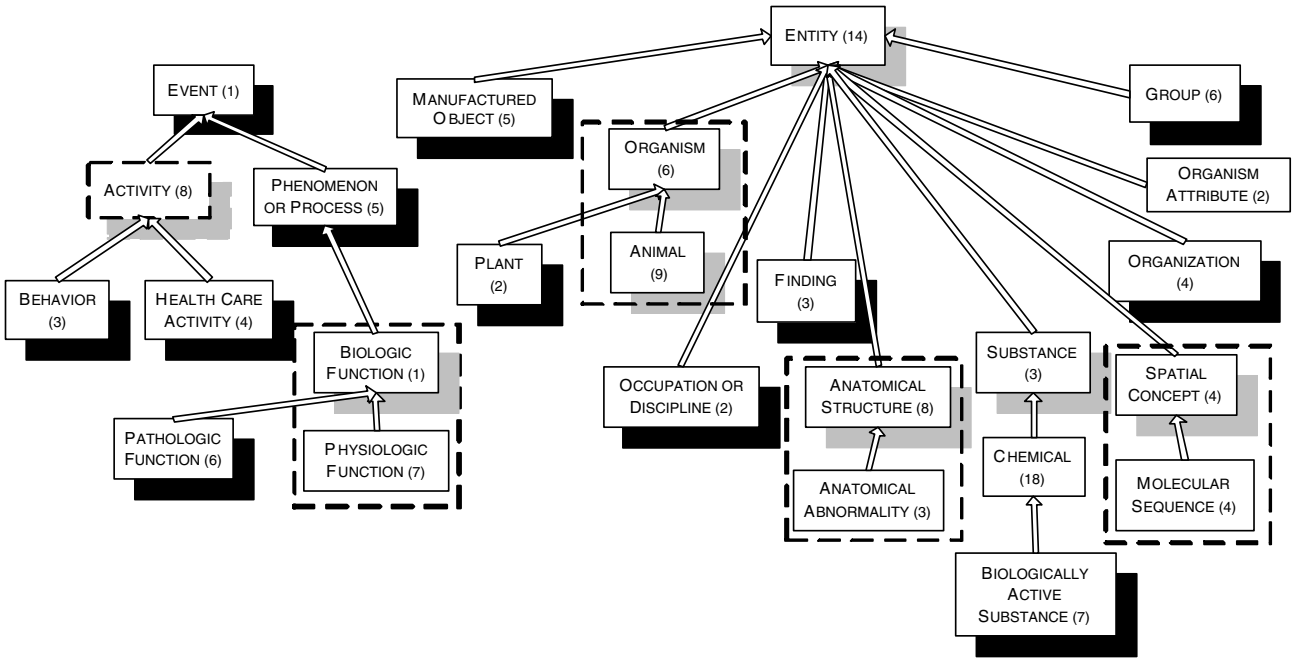
Once M^*_B and SD are obtained, M''_B can be derived. It contains all MSTs from SD and M^*_B, CHEMICAL(8), PHARMACOLOGIC SUBSTANCE(2) and ORGANIC CHEMICAL(8), as well as

Table 11
Weight distribution of the top-down and bottom-up metaschemas

Weight	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Top-down	3	4	3	2	3	3	1	3	0	0	1	0	0	0	1	0	0	0
Bottom-up	1	3	4	4	2	3	7	2	1	0	0	0	0	1	0	0	0	1



(a) Top-down Consensus metaschema: 26 MSTs



(b) Bottom-up Consensus metaschema: 25 MSTs

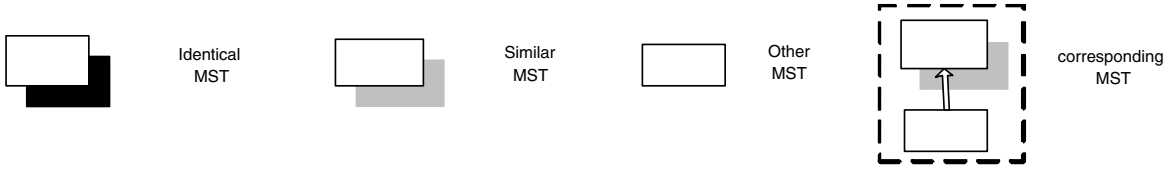


Fig. 13. The consolidation of corresponding MSTs.

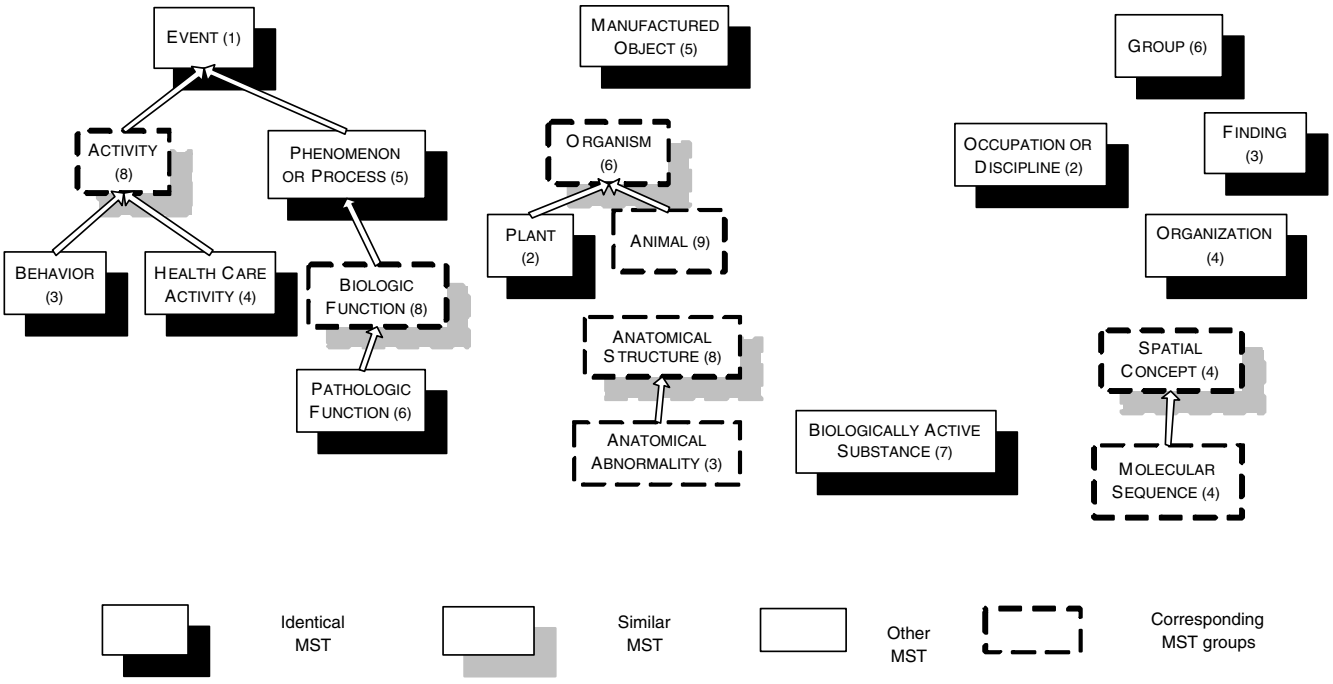


Fig. 14. The partial consolidated metaschema with corresponding MSTs.

the *meta-child-of* relationships from the last two MSTs to CHEMICAL(8).

4.4.13. Step 13 for **Chemical**

Comparing M''_A, which consists of CHEMICAL(18), of the bottom-up metaschema and M''_B, consisting of CHEMICAL(8), PHARMACOLOGIC SUBSTANCE(2) and ORGANIC CHEMICAL(8), we see that this is a case of similar MSTs with a refinement. This situation is handled as discussed in Step 3. The refinement {CHEMICAL(8), PHARMACOLOGIC SUBSTANCE(2), ORGANIC CHEMICAL(8)} is selected for the consolidated metaschema rather than CHEMICAL(18) because it contributes less to the standard deviation, compared to CHEMICAL(18). Note that the only MST of the consolidated metaschema (of the SUBSTANCE component) which is not an MST of a given metaschema is CHEMICAL(8).

4.4.14. Step 5 for **Entity**

For the subtree rooted at **Entity**, ENTITY(1) is less expensive than ENTITY(14) (see Fig. 16), in terms of the standard deviation contribution. Therefore ENTITY(1) in the top-down metaschema is chosen (see Fig. 16a).

4.4.15. Step 6 for **Entity**

ENTITY(1) will be removed from the top-down metaschema together with the two *meta-child-of* relationships directed to it, as shown in Fig. 16a. An M''_A, which is {PHYSICAL OBJECT(1), CONCEPTUAL ENTITY(5), INTELLECTUAL PRODUCT(3), IDEA OR CONCEPT(6)} (Fig. 17a), is obtained as a result of the deletion. The metaschema M''_A consists of two disconnected MST subtrees, one rooted at PHYSICAL OBJECT(1) and the other at CONCEPTUAL ENTITY(3).

4.4.16. Step 7 for **Physical Object and Conceptual Entity**

The expanded Semantic Network SN″ constructed from M''_A, consists of two subtrees rooted at **Physical Object** and **Conceptual Entity**, respectively (Fig. 17b).

4.4.17. Summary of Steps 8–13

Since ENTITY(1) is not an MST of the bottom-up metaschema, an M^*_B needs to be derived, which consists of ORGANISM ATTRIBUTES(2) only (see Fig. 17c). The Semantic Network SN* constructed from M^*_B includes **Organism Attribute** and **Clinical Attribute** (Fig. 17d). We can then obtain SD as {PHYSICAL OBJECT(1), CONCEPTUAL ENTITY(12)} as a grouping of the difference D of SN″ and SN* into maximally connected components (Fig. 17e). The resulting M''_B combining M^*_B and SD and all *meta-child-of* links connecting them is therefore {PHYSICAL OBJECT(1), CONCEPTUAL ENTITY(12), ORGANISM ATTRIBUTE(2)} (Fig. 17f). Since M''_A and M''_B are not empty, we apply the R-CONSOLIDATE procedure recursively to the M''_A of the top-down metaschema and the M''_B of the bottom-up metaschema and their common Semantic Network SN″. For space reasons, we omit the details of this recursive call.

4.4.18. Obtaining completely consolidated metaschema M_C

In the modification stage of the MAIN-CONSOLIDATE algorithm, we scan the entire consolidated metaschema, looking for singleton MSTs which are not

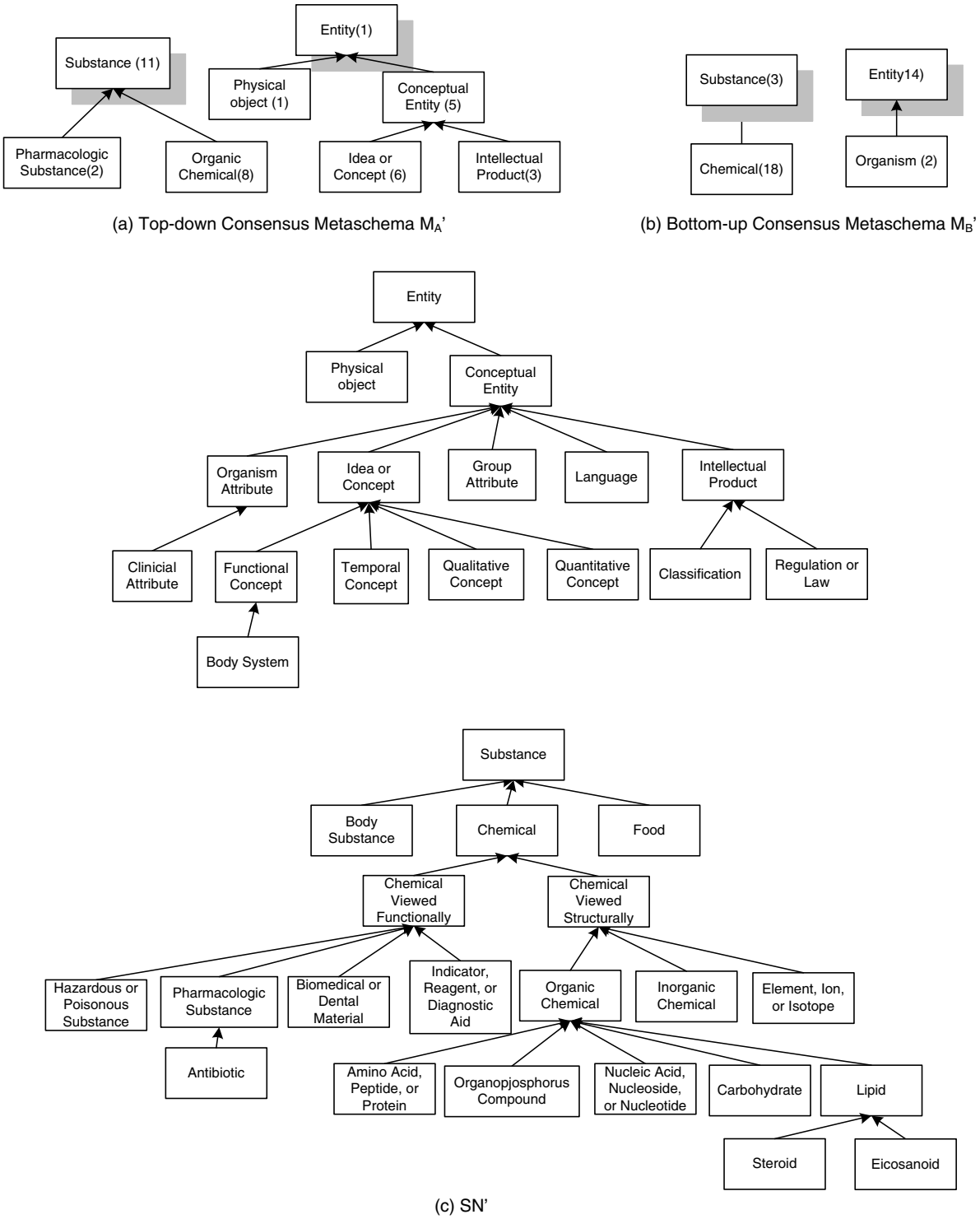


Fig. 15. Auxiliary metaschemas M'_A, M'_B and expanded Semantic Network SN′.

branching points in the metaschema. We see that the singleton MST PHYSICAL OBJECT(1) is not a branching point in the metaschema. Therefore, PHYSICAL OBJECT(1) is merged with its parent ENTITY(1) to create the MST ENTITY(2). The MST MANUFACTURED OBJECT(5) which was *meta-child-of* PHYSICAL OBJECT(1), is now *meta-child-of* ENTITY(2).

4.5. Final consolidated metaschema

The consolidated metaschema is shown in Fig. 18. There are 28 MSTs in the consolidated metaschema. Twelve MSTs come from the identical MSTs in the two original consensus metaschemas. Eight MSTs are taken from cases of the refinements, seven of which come from the bottom-

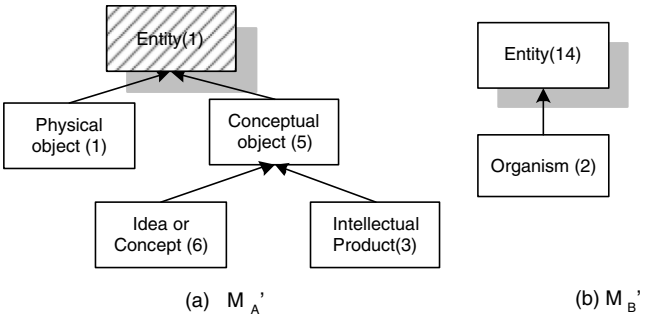


Fig. 16. M'_A and M'_B for **Entity**.

up consensus metaschema. Only BIOLOGICAL FUNCTION(8) comes from the top-down consensus metaschema. The remaining 8 MSTs come from the two subtrees rooted in non-refinable similar MSTs, namely **Entity** and **Substance**. Among those 8 MSTs, CONCEPTUAL ENTITY(5), INTELLECTUAL PRODUCT(3), IDEA OR CONCEPT(6), ORGANIC CHEMICAL(8) and PHARMACOLOGIC SUBSTANCE(2) are from the top-down consensus metaschema, while only SUBSTANCE(3) is from the bottom-up consensus metaschema. ENTITY(2) and CHEMICAL(8) are the only two MSTs, of the consolidated metaschema which do not appear as MSTs in the top-down or bottom-up metaschemas. They are still similar to MSTs, in the two given metaschemas, respectively, but with different groups. The structural properties of the consolidated metaschema will be compared to those of the consensus metaschemas in Section 5.

We now turn our attention to non-IS-A, i.e., associative relationships. An occurrence of an associative relationship at an MST is called an introduction occurrence for this

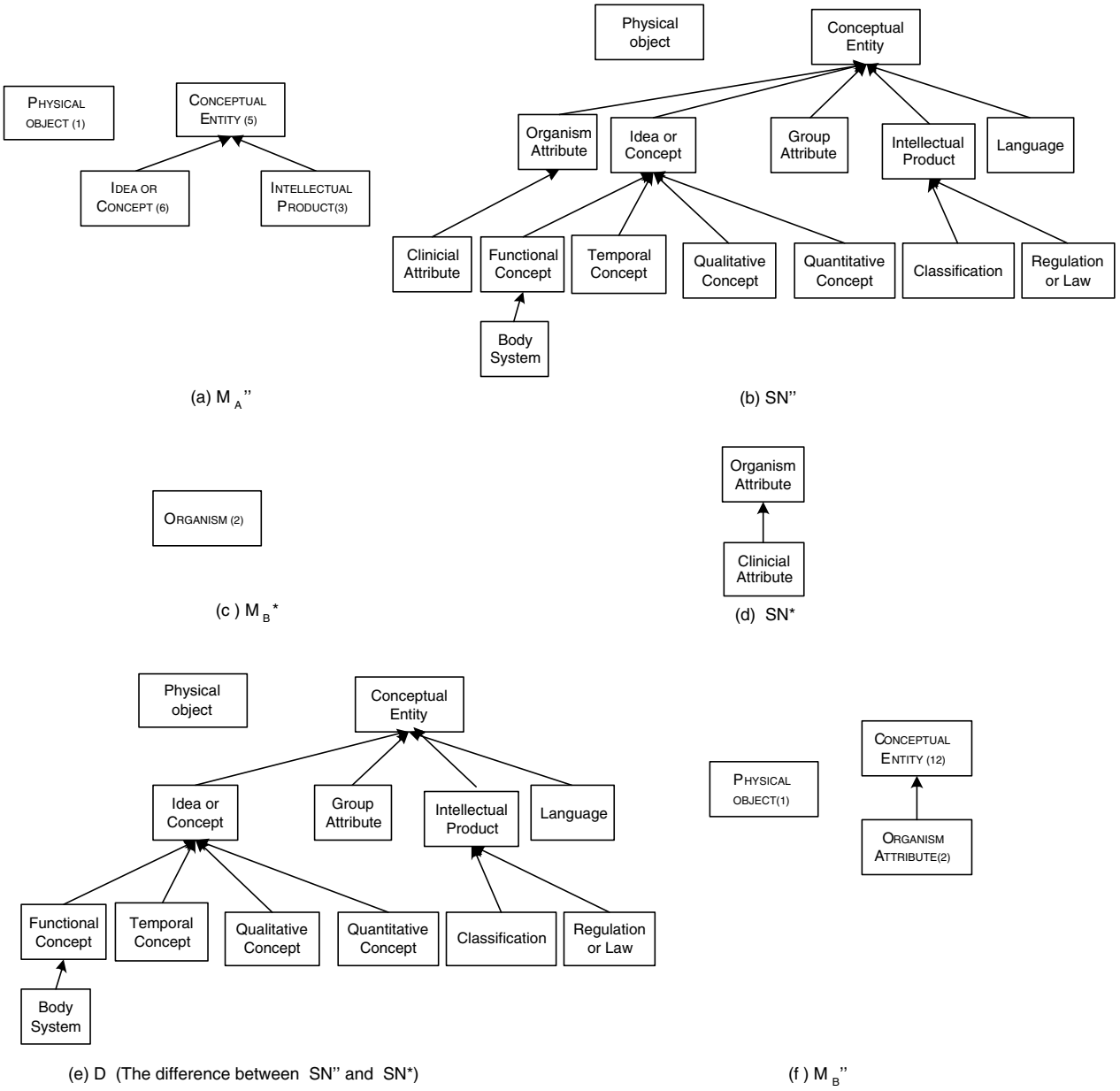


Fig. 17. Processing for **Entity**.

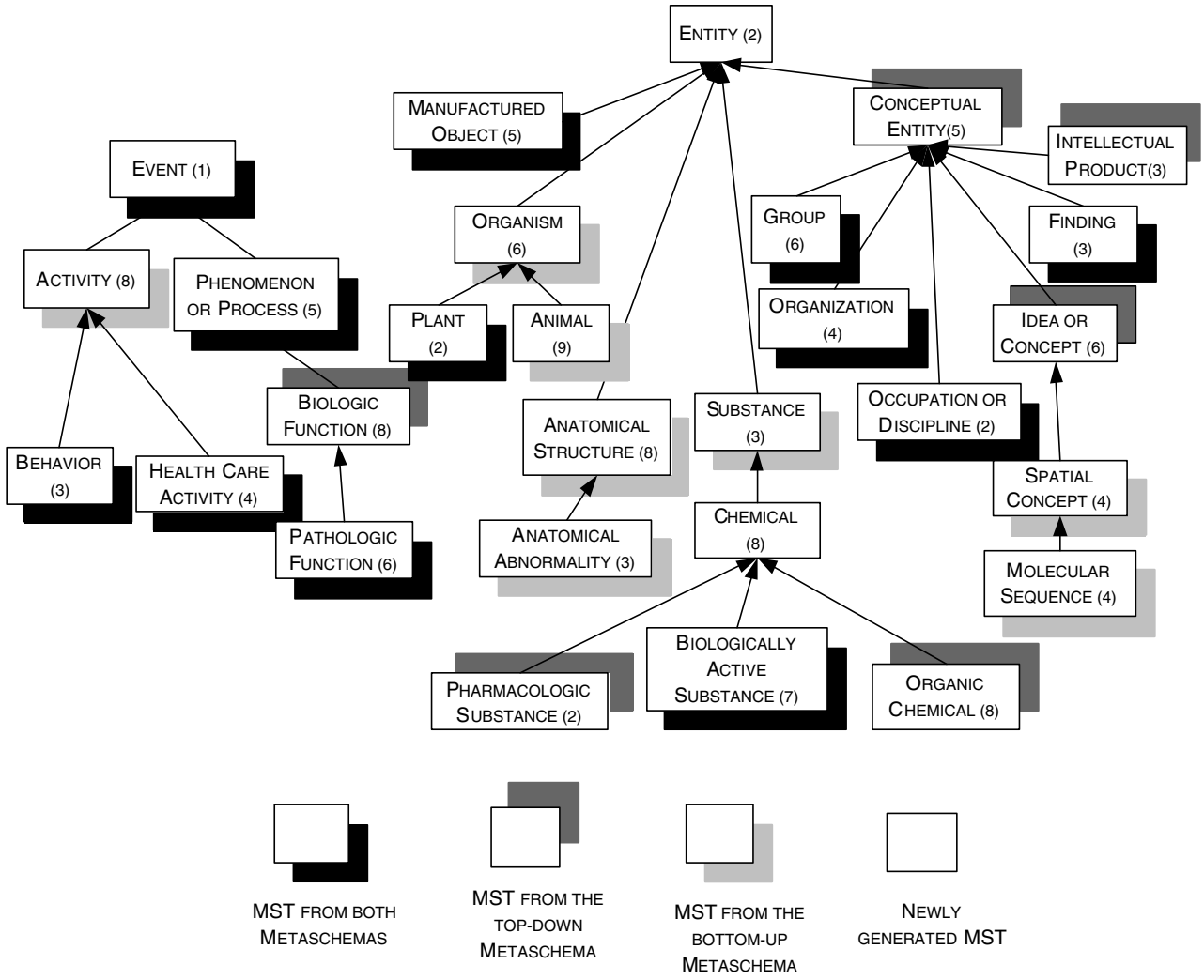


Fig. 18. Consolidated metaschema hierarchy.

kind of relationship if this kind of relationship is not defined for the parent of this MST. We note that the same kind of relationship, e.g. "issue in" may have several introduction occurrences at several MSTs, none of which is an ancestor of another. All other occurrences of this kind of relationship are inherited from some introduction occurrence of this kind of relationships. The complete presentation of the consolidated metaschema including the introduction occurrences of associative (semantic) relationships appears in Fig. 19. For space reasons, we had to code each relationship name as a number (see Table 12). The inherited relationships are omitted from Fig. 19 to reduce its graphical complexity, as they can be deduced due to the inheritance. However, the inherited associative relationships are taken into account in calculating the complexity of the metaschema (see Section 5).

5. Discussion

The purpose of our research was to obtain a high quality metaschema. As discussed in the background section, the motivation for conducting the bottom-up metaschema study came from our dissatisfaction with the large variations in the top-down experts' metaschemas. We hoped that the bottom-up heuristics would lead experts to produce metaschemas with less variability and higher agreement due to the higher amount of data considered in the bottom-up process compared with the top-down approach. The results indeed show lower variations among the experts (Table 13). Note that although the average inter-participant agreement is higher for the top-down study (16.76 vs. 14.41), this is misleading due to larger cardinalities for the top-down study. When the average agreement is measured relative to the average cardinality (26.73 vs. 23), the higher proportion of agreement is obtained for the bottom-up study (0.67 vs. 0.63).

Although the variability of the experts' metaschemas is lower in the bottom-up study than in the top-down study, it is still quite high. The large range shows the high variability of participant responses. Thus, there is a problem when using any individual expert's metaschema to evaluate an

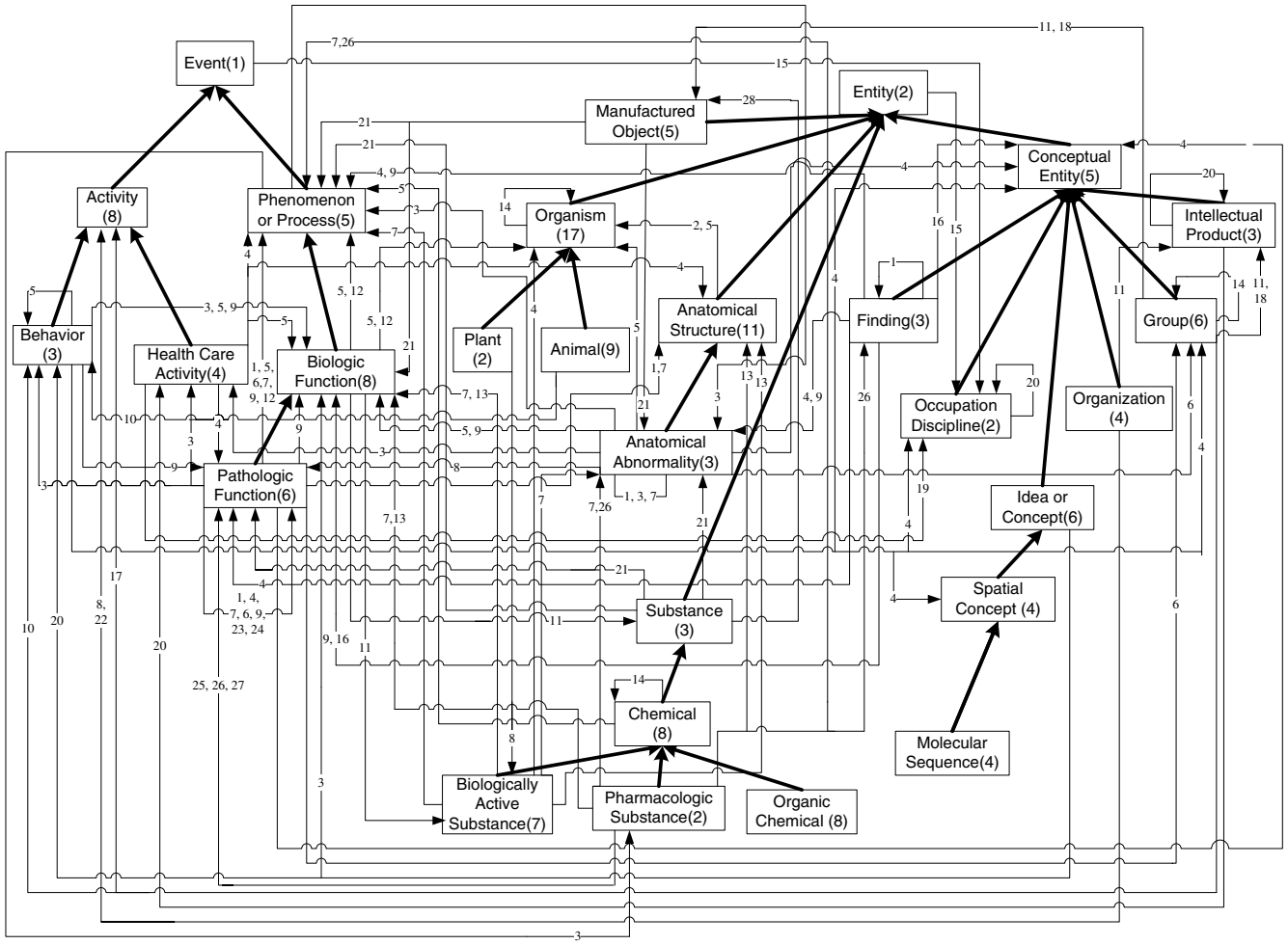


Fig. 19. Consolidated metaschema with relationships.

Table 12
Relationship number codes

Number	Relationship	Number	Relationship	Number	Relationship
1	co-occurs_with	2	part_of	3	result_of
4	associated_with	5	affects	6	occurs_in
7	complicates	8	location_of	9	manifestation_of
10	exhibits	11	produces	12	process_of
13	disrupts	14	interacts_with	15	issue_in
16	evaluation_of	17	performs	18	uses
19	method_of	20	conceptual_part_of	21	causes
22	carries_out	23	precedes	24	degree_of
25	diagnoses_of	26	treats	27	prevents
28	ingredient_of				

algorithmically derived metaschema, as experts vary so much in their opinions (Section 2).

It seems that one cannot rely on any one expert to provide an authoritative metaschema for SN. At first this sounds quite disappointing. However, this phenomenon is understandable when one realizes that there is an exponential number of connected partitions for the SN, each of which would lead to a different metaschema. Furthermore, the experts are asked to make choices of importance and distinctions which are subjective and are influenced by their experience, background, specialty and personal preferences. Therefore we derived the consensus metaschemas for both studies, to overcome the variability of the individual experts' metaschemas.

The two consensus metaschemas, the top-down metaschema and the bottom-up metaschema, reflect human considerations, since they are metaschemas resulting from several human experts' input. At the same time, their cumulative nature helped to overcome the variability mentioned above. However, as their evaluations show, each of

Table 13
Comparing experts' metaschemas of both studies

Experts' metaschema properties	Top-down study	Bottom-up study
Lowest cardinality	12	16
Highest cardinality	36	30
Cardinality range	25	15
Average cardinality	26.73	23
Cardinality standard deviation	10.23	4.56
Upper threshold	45	41
Lower threshold	2	3
Threshold range	44	39
Average inter-participant agreement	16.76	14.41
Ratio of average agreement to average cardinality	0.63	0.67

them has pros and cons, in term of its structural properties. Thus, we constructed the consolidated metaschema that can best facilitate user orientation into the SN, by enjoying the advantages of each of the consensus metaschemas and avoiding their disadvantages.

Altogether, in the consolidated metaschema, the groups identical to both input metaschemas contain 48 (36%) STs, the groups identical only to those in the bottom-up consensus metaschema contain 45 (33%) STs and the groups identical only to the top-down consensus metaschema contain 32 (24%) STs. Only 10 (7%) STs are contained in MSTs which did not appear in the given metaschemas, although they are similar to such MSTs (according to the definition of similar MSTs in Section 3.2). Hence, in the consolidated metaschema, 93% of the STs from the SN appear in MSTs selected by human experts. In [10], McCray et al. define the naturalness property of a partition of the SN as the condition that "the groups must characterize the domain in a way that is acceptable to a domain expert." As we saw, all MSTs except for two satisfy the naturalness property, as required for a partition of the SN by McCray et al. [10].

From the semantic viewpoint, we can see that important MSTs, missing in either of the consensus metaschemas but occurring in the other, are now in the consolidated metaschema. For example MOLECULAR SEQUENCE(4), ANATOMICAL ABNORMALITY(3) and ANIMAL(9) are missing from the top-down consensus metaschema. Similarly, ORGANIC CHEMICAL(8), INTELLECTUAL PRODUCT(3), IDEA OR CONCEPT(6), PHARMACOLOGIC SUBSTANCE(2) and CONCEPTUAL ENTITY(5) are missing from the bottom-up consensus metaschema. All these MSTs appear in the consolidated metaschema. In addition, the consolidated metaschema shows better structural properties than either of the input metaschema (see Table 14). In particular, the consolidated metaschema has a lower weight range and a lower standard deviation.

There were no statistically significant differences of the average weights among the top-down, bottom-up and consolidated metaschemas.

By an F test for equality of variance [35], the standard deviations of the top-down and bottom-up metaschemas did not differ, but both were greater than that of the consolidated metaschema ($p < .05$). Using a bootstrap estimator [36], the maximum weight of the top-down metaschema was determined to be statistically significantly greater than that of the consolidated metaschema, but the other differences among the maxima did not achieve statistical significance.

Table 14
Comparison of the structural measures for the top-down, bottom-up and consolidated metaschemas

Measures	Top-down	Bottom-up	Consolidated
Cardinality	26	25	28
Complexity	4.3	4.72	3.92
Maximum weight	15	18	9
Minimum weight	1	1	1
Weight spread	14	17	8
Average weight	5.19	5.40	4.89
Standard deviation	3.49	3.93	2.25
Coverage (%)	70.6	75.93	75.05

As can be seen, the bottom-up consensus metaschema contributed more of its groups to the consolidated metaschema than the top-down metaschema. On one hand, this is in line with the bottom-up study being based on providing the experts with more data and the lower variability of this study. On the other hand, this is surprising in view of (Table 13) better structural measures of the top-down consensus metaschema with regard to maximum weight and standard deviation. However, as mentioned before, these differences are not statistically significant. In spite of the much lower variability of the bottom-up experts' metaschemas, the advantage of the bottom-up study disappeared when we obtained the consensus metaschemas. We attribute this phenomenon to the largest MST, CHEMICAL(18) in the bottom-up consensus metaschema. Working bottom-up, most of the experts did not identify any ST in the group CHEMICAL(18), while experts scanning the SN top-down, identified ORGANIC CHEMICAL(8) and PHARMACOLOGIC SUBSTANCE(2). This largest weight contributed much more to the weight range and standard deviation than the largest MST ORGANISM(15) in the top-down consensus metaschema. However, this does not imply better quality of the top-down consensus metaschema for the other MSTs. Interestingly, the MAIN-CONSOLIDATE algorithm avoided selecting any large MSTs, and instead chose smaller natural MSTs covering the STs of the large MSTs in the other metaschema.

6. Conclusions

In this paper, we compared, from a structural point of view, two heuristic metaschemas, the top-down and bottom-up consensus metaschemas, derived from two studies involving two groups of UMLS experts. We defined different levels of similarity. Using these definitions, we found that both heuristic metaschemas agree in almost half (12) of the MSTs. There are seven similar MSTs, five of which are cases of refinement. We defined several structural properties of a metaschema, such as, cardinality, complexity, maximum weight, minimum weight, weight spread, average

weight, standard deviation and coverage percentage. We used them for evaluation of both consensus metaschemas. Second, we designed an algorithm to construct a consolidated metaschema from two given metaschemas, enjoying the advantages and avoiding the disadvantages of both. The consolidated metaschema has better structural properties, such as lower weight range as well as lower standard deviation than its inputs. It can better serve as an abstract network and support user orientation and navigation of the Semantic Network, due to its naturalness in identifying many groups selected by most experts in either of the studies. Furthermore, the relative structural uniformity of the consolidated metaschema, as expressed in the low weight range and standard deviation, will also support user orientation and navigation when accessing the underlying ST groups using various SN graphical views described in [12,15].

In previous research we had derived algorithmically generated metaschemas. The consolidated metaschema, being a digest of many domain experts' input, can be used to evaluate the naturalness of the algorithmic cohesive [12] and lexical [15] metaschemas, rather than the top-down consensus metaschema used for their evaluation in [12] and [16], respectively.

Acknowledgments

We thank the following researchers who have published on the UMLS or related topics for participating in any of the reported studies. The list is in alphabetical order. We apologize for possible inaccuracies caused by some respondents sending their responses in unnamed self-addressed envelopes: G. Octo Barnett, Charles Barr, James Brinkley, Christopher G. Chute, James J. Cimino, Gregory F. Cooper, Gai Elhanan, Mark Erlbaum, Carol Friedman, Robert A. Greenes, Stephen Johnson, Kenric Hammond, William R. Hersh, Pat Malholt, Jose L. Mejino, Perry L. Miller, Randolph Miller, Nels E. Olsen, Cornelius Rosse, Barry Smith and Kent A. Spackman.

References

[1] Tuttle MS, Sherertz DD, Olson NE, Erlbaum MS, Sperzel WD, Fuller LF et al. Using META-1, the first version of the UMLS Metathesaurus. In: Proc. fourteenth annual SCAMC; 1990. p. 131–5.

[2] Schuyler PL, Hole WT, Tuttle MS, Sheretz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. Bull Med Libr Assoc 1993;81(2):217–22.

[3] McCray AT. UMLS Semantic Network. In: Proc. thirteenth annual SCAMC. Washington, DC; 1989. p. 503–7.

[4] McCray AT, Hole WT. The scope and structure of the first version of the UMLS Semantic Network. In: Proc. fourteenth annual SCAMC, November 1990. Los Alamitos, CA; 1990. p. 126–30.

[5] McCray AT. Representing biomedical knowledge in the UMLS Semantic Network. In: Broering NC, editor. High-performance medical libraries: advances in information management for the virtual era. Westport, CT: Mekler; 1993. p. 45–55.

[6] McCray AT. An upper-level ontology for the biomedical domain. Comp Func Genom 2003;4:80–4.

[7] Campbell KE, Oliver DE, Shortliffe EH. The Unified Medical Language System: toward a collaborative approach for solving terminologic problems. JAMIA 1998;5(1):12–6.

[8] Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. JAMIA 1998;5(1):1–11.

[9] U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine. Unified Medical Language System (UMLS); 2006. Available from: www.nlm.nih.gov/research/umls.

[10] McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. In: Proc. medinfo, September 2001. London, UK; 2001. p. 171–5.

[11] Kumar A, Smith B, Stefanelli M, Quaglinia S, Piazza M. Implementing clinical guidelines in an organizational setup. Proceedings of the workshop on model-based and qualitative reasoning in biomedicine, p. 39–44.

[12] Perl Y, Chen Z, Halper M, Geller J, Zhang L, Peng Y. The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network. J Biomed Inform 2003;35(3):194–212.

[13] Zhang L, Perl Y, Halper M, Geller J. Designing metaschemas for the UMLS Enriched Semantic Network. J Biomed Inform 2003;36(6):433–49.

[14] Zhang L, Perl Y, Geller J, Halper M, Cimino JJ. An enriched UMLS Semantic Network with a multiple inheritance hierarchy. J Am Med Inform Assoc 2004;11(3):195–206.

[15] Zhang L, Perl Y, Halper M, Geller J, Hripcsak G. A lexical metaschema for the UMLS Semantic Network. Artif Intell Med 2005;33:41–59.

[16] Zhang L, Hripcsak G, Perl Y, Halper M, Geller J. An expert study evaluating the UMLS lexical metaschema. Artif Intell Med 2005;34(3):219–33.

[17] Gu H, Perl Y, Elhanan G, Min H, Zhang L, Peng Y. Auditing concept categorizations in the UMLS. Artif Intell Med 2004;31(1):29–44.

[18] Niles I, Pease A. Origins of the IEEE standard upper ontology. In: Working notes of the IJCAI-2001 workshop on the IEEE standard upper ontology, August 2001. Seattle, WA; 2001.

[19] Niles I, Pease A. Towards a standard upper ontology. In: Proc. FOIS, October 2001. Ogunquit, MA; 2001.

[20] Fellbaum C, editor. WordNet: an electronic lexical database (language, speech, and communication). The MIT Press; 1998.

[21] Niles I, Pease A. Linking lexicons and ontologies: mapping wordnet to the suggested upper merged ontology. In: Proc. international conference on information and knowledge engineering, June 2003 (IKE'03). Las Vegas, NV; 2003.

[22] Gu H, Halper M, Geller J, Perl Y. Benefits of an object-oriented database representation for controlled medical terminologies. JAMIA 1999;6(4):283–303.

[23] Cimino JJ, Clayton PD, Hripcsack GJ, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. JAMIA 1994;1(1):35–50.

[24] Liu L, Halper M, Geller J, Perl Y. Controlled vocabularies in OODBs: modeling issues and implementation. Distrib Parallel Databases 1999;7(1):37–65.

[25] Gu H, Perl Y, Halper M, Geller J, Kuo F, Cimino JJ. Partitioning an object-oriented terminology schema. Methods Med Inform 2001;40:204–12.

[26] Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. J Biomed Inform 2003;36(6):414–32.

[27] Nelson SJ, Sheretz DD, Tuttle MS, Erlbaum MS. Using MetaCard: a hypercard browser for biomedical knowledge sources. In: Proc. annu symp comput appl med care; 1990. p. 151–4.

[28] Li Q, Shilane P, Noy NF, Musen MA. Ontology acquisition from online knowledge sources. In: Proc. 2000 AMIA annual symposium, November 2000. Los Angeles, CA; 2000. p. 497–501.

[29] Tuttle MS, Cole WG, Sheretz DD, Nelson SJ. Navigating to knowledge. Methods Inf Med 1995;34(1–2):214–31.

[30] Aristotle. The categories. Cambridge, MA: Harvard University Press; 1973.
[31] Hripcsak G, Friedman C, Anderson P, DuMouchel W, Johnson S, Clayton P. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995;122:681–8.
[32] Dunn G. Design and analysis of reliability studies. New York: Oxford University Press; 1989.
[33] Perl Y, Schach SR. Max–min tree partitioning. JACM 1981;28(1):5–15.
[34] Perl Y, Becker R, Schach S. A shifting algorithm for min max tree partitioning. JACM 1982;29:58–76.
[35] Snedecor GW, Cochran WG. Statistical methods. 8th ed. Ames, IA: Iowa State University Press; 1989.
[36] Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993.

Glossary of metaschemas

Metaschema: A compact, acyclic Abstraction Network of the SN based on a partition of the IS-A hierarchy of the SN into connected components. The nodes of the metaschema, called metasemantic types (MSTs), represent connected groups of semantic types of the SN. The MSTs are linked by hierarchical *meta-child-of* relationships and non-hierarchical *meta-relationships*.

Cohesive metaschema: A metaschema created by an algorithm based on structural considerations, described in [12].

Lexical metaschema: A metaschema created by an algorithm based on lexical considerations, described in [15].

Expert metaschema: A metaschema created by a domain expert.

Top-down expert metaschema: An expert metaschema created using a heuristic methodology, processing the SN starting at its roots and proceeding down the tree. It is described in [16] and reviewed in Section 2.2.

Bottom-up expert metaschema: An expert metaschema created using a heuristic methodology processing the SN starting at its leaves and proceeding up the tree. It is described in Section 3.1.

Cumulative metaschema: A metaschema resulting from aggregating experts' metaschemas according to a threshold value.

Consensus metaschema: The cumulative metaschema resulting from aggregating the experts' metaschemas according to a threshold of a simple majority of experts.

Top-down consensus metaschema: A consensus metaschema of the top-down expert metaschemas.

Bottom-up consensus metaschema: A consensus metaschema of the bottom-up expert metaschemas.

Consolidated metaschema: A metaschema created by the MAIN-CONSOLIDATE algorithm, combining the best features of the top-down consensus metaschema and the bottom-up consensus metaschema. This is the main result of this paper.